DOCUMENT RESUME

ED 060 658		24	EM 009 708

AUTHOR Paulus, Dieter H.; And Others
TITLE Computer Simulation of Human Ratings of Creativity. Final Report.
INSTITUTION Connecticut Univ., Storrs. School of Education.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO BR-9-A-032
PUB DATE Nov 70
GRANT OEG-1-9-090032-0108(010)
NOTE 92p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Computers; Content Analysis; Creative Ability; *Creativity; *Creativity Tests; Multiple Regression Analysis; Predictor Variables; Reliability; *Scoring; *Simulation; Validity; Verbal Ability
IDENTIFIERS *Torrance Tests of Creative Thinking; TTCT

ABSTRACT

The Torrance Tests of Creative Thinking (TTCT) may represent a breakthrough in the area of creativity research, since they provide a functional instrument for measuring creative potential in children, adolescents, and adults. However, there are certain technical problems with achieving interscorer reliability which may act as a deterrent to the widespread use of the tests. A thorough review of the literature has revealed that computerized content analysis has not been used for scoring creativity tests, although it appears to be an appropriate approach. Therefore, a study was devised to develop strategies appropriate to computer scoring of the TTCT, to determine the effectiveness of actuarial measures in the prediction of scores, and to make some initial explorations regarding the appropriateness of the norms developed by Torrance for scoring the tests. Responses of 153 subjects to Verbal Form A of the tests were reliably rated by four trained judges, and multiple correlation analyses were computed. The regression equations generated through this process were shown to have high predictive power, and examination of the important predictors suggested that the TTCT has a single underlying dimension of verbal fluency. Category count variables were useful in the prediction process. (Author/SH)

Final Report

Project No. 9-A-032
Grant No. OEG-1-9-090032-0108(010)

Dieter H. Paulus
Joseph S. Renzulli
Francis X. Archambault, Jr.

School of Education
University of Connecticut
Storrs, Connecticut 06268

Computer Simulation of Human Ratings of Creativity

NOVEMBER 1970

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development
(Regional Research Program)

FINAL REPORT

Project No. 9-A-032

Grant No. OEG-1-9-090032-0108(010)

COMPUTER SIMULATION OF HUMAN RATINGS OF CREATIVITY

Dieter H. Paulus
University of Connecticut

Joseph S. Renzulli
University of Connecticut

Francis X. Archambault, Jr.
Boston University

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects and Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

Final Report

Project No. 9-A-032
Grant No. OEG-1-9-090032-0108(010)

Dieter H. Paulus
Joseph S. Renzulli
Francis X. Archambault
The University of Connecticut
School of Education
Storrs, Connecticut 06268

COMPUTER SIMULATION OF HUMAN RATINGS OF CREATIVITY

November, 1971

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development

(Regional Research Program)

ABSTRACT

The purposes of this project were 1) to develop strategies appropriate to computer scoring the Torrance Tests of Creative Thinking, 2) to determine the effectiveness of actuarial measures in the prediction of scores assigned to these tests and 3) to make initial explorations regarding the appropriateness of the norms developed by Torrance for scoring these tests. A review of the literature showed that no previous attempt has been made to apply a computerized content analytic technique to the scoring of creativity tests but that such an attempt is appropriate. Responses of 153 subjects to Verbal Form A were reliably rated by four trained judges. Multiple correlation analyses involving predictor variables (including "variables of opportunity" and variables based on the Torrance norms) and the criterion variables of Fluency, Flexibility and Originality for the various activities were computed, and regression equations were generated. Through cross-validation, these equations were shown to have high predictive power. Examination of the nature of the important predictors suggested the possibility that the TTCT has a single underlying dimension - verbal fluency. Category count variables based on Flexibility and Originality dictionaries were useful in the prediction process.

PREFACE

The findings reported herein are the result of a cooperative effort on the part of several persons who have made direct and indirect contributions to this study. Because of the longstanding interest in natural language computing at The University of Connecticut the writers have been able to draw upon many resources and the research competencies of a number of persons who have worked in this area of study. The work of these persons is cited in the report that follows. We would like to express our sincere appreciation to each of the individuals whose previous and sometimes pioneering work in the area of natural language computing has helped to make our work somewhat more manageable.

We are especially indebted to two former University of Connecticut researchers who have made direct contributions to the study. Dr. John F. Greene of the University of Bridgeport was responsible for developing computer programs and processing a major segment of the data related to the study. Chapter V of this report was prepared by Dr. Greene and his work also is reflected in the discussion of findings that is reported in Chapter VI. Dr. Gerald A. Fisher, presently at the Illinois Institute of Technology, was responsible for developing the SCORTXT program which constituted the major programming system used in the study.

The writers are indebted to Dr. E. Paul Torrance of the University of Georgia and Dr. Robert N. Walker of Personnel Press, Inc. who assisted us in many ways throughout the course of the project and who provided us with constant encouragement in this relatively new area of research endeavor.

We are grateful to Carolyn Callahan who supervised the scoring of the Torrance Tests of Creative Thinking and to the following persons who served as test scorers on the project: Margaret Capellini, Tobi Okun Cannelli, Jane Goodman, Michael Lunney, Christine deRenne Stephen, and Sandra Wever-Bey.

We are indebted to a number of persons who generously provided us with data for a pilot study that paved the way for the present investigation. These persons include: Dr. John R. Stubbings of the Alexandria, Virginia Public Schools, Dr. Donald J. Treffinger of Purdue University, Dr. Richard E. Ripple of Cornell University, and Dr. John S. Dacey of Boston College. We also are indebted to Dr. Hugh Clark, Associate Dean of the Graduate School for Research at The University of Connecticut and Dr. Richard V. McCann, Director of Research for Region I of the United States Office of Education who offered us many valuable suggestions in the preparation of the proposal for this research.

Finally, we would like to express our thanks to The University of Connecticut Computer Center for their cooperation in processing our data and to the many students and their teachers who participated in the study.

Dieter H. Paulus
Joseph S. Renzulli
Francis X. Archambault
Storrs, Connecticut
November, 1970

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLES

TABLES (Continued)

TABLES (Continued)

FIGURES

CHAPTER I

STATEMENT OF THE PROBLEM

Since the last decade when Guilford (1950) called attention to the virtual neglect of the concept of creativity by American researchers, there has been an enormous expansion of interest and research in the nature of this higher mental process. A myriad of problems and controversies have surrounded work in the area of creativity, but one of the most pressing issues continually has been the search for valid and reliable means of measuring creative performance.

Workers in both the laboratory and the schools have recognized the need for valid and reliable devices to assess creative potential, but it is in the schools that the need is most acute. Curriculum specialists and classroom teachers alike are being asked to make provisions for highly creative youngsters, and provisions are being made. But the students who benefit from these programs are being selected by a variety of subjective processes without the aid of a reliable identification instrument. Paulus and Renzulli (1969) have commented on this unfortunate state of affairs: "In other words, if we cannot accurately and economically determine who our most potentially creative youngsters are, then efforts to 'do something' for the highly creative are analogous to prescribing medicine before diagnosing the illness."

The recent publication of the _Torrance Tests of Creative Thinking_ (Torrance, 1966a) in many respects may be regarded as a breakthrough in the area of creativity measurement. Based on nearly nine years of research and development by Torrance and his colleagues, the tests represent a pioneering venture in that they provide the researcher and educational practitioner with a _functional_ instrument for measuring creative potential in children, adolescents, and adults. In spirt of the relatively high level of development of the Torrance instruments, certain technical problems related to levels of training on the part of scorers may act as deterrents to their widespread use. Torrance (1966c) reports correlation coefficients of interscorer reliability ranging from .76 to .98 for the different sub-tests; however, he points out the possibility of errors in scoring that exist when scorers fail to read carefully the scoring guide or to scan adequately the weights assigned to certain dimensions of creativity. At least one reviewer (Hoepfner, 1967), has called attention to this problem and also has suggested that the time required to score the test battery may be a relatively long affair.

Since the advent of the General Inquirer (Stone, _et al._, 1966) and other strategies, computerized content analysis has been effectively used in the solution of many research problems. However, a thorough review of the literature has revealed no previous research applying machine strategies to the scoring of creativity tests. That the approach is appropriate is attested to, nevertheless, by successful applications in related areas of natural language processing (Hiller, 1967; Page and Paulus, 1968; McManus, 1968; Marcotte, 1969).

The purposes of the present research are: 1) to develop strategies appropriate to computer scoring the Torrance Tests of Creative Thinking (1966a); 2) to determine the effectiveness of actuarial measures, such as average sentence length, in the prediction of scores assigned to these tests; and 3) to make some initial explorations regarding the appropriateness of the norms developed by Torrance (1966b) for scoring these tests.

The remainder of this work is divided into five chapters. The most pertinent related literature for both content analysis and creativity is presented in Chapter II. Chapter III discusses the procedures followed in each of the various phases of the research. In Chapters IV and V the results of this research are presented in tabular and narrative form. A discussion of these results and their implications as well as some theoretical considerations for future research are given in Chapter VI.

CHAPTER II

RELATED LITERATURE

Introduction

Since the automated scoring of responses to the Torrance Tests of Creative Thinking (1966a), is heavily dependent upon the strategies of content analysis, a review of the literature dealing with various aspects of analyzing the content of written material will be presented here. A general review of relevant creativity research also will be given.

Content Analysis

The lable "content analysis" has been applied to many widely divergent research methods which, as Zieky (1968, p. 16) has observed, "have little in common other than the fundamental underlying assumption that there exist certain elements in any given communication which may be utilized as indicators from which inferences might be made, or by means of which the communication itself might be described." Most recent formal definitions, however, have tended to restrict the applications of the term solely to those studies which attempt the analysis of communication content in accordance with generally accepted criteria of sound scientific research (Berelson, 1952; Cartwright, 1953; Stone, Dunphy, Smith and Ogilvie, 1966). A concise, yet representative definition is that proposed by Holstoi (1966, p. 10): "Content analysis is any technique for making inferences by systematically and objectively identifying specified characteristics of messages."

Operationally, content analysis consists of the reduction of a text to those preselected characteristics which the researcher feels are relevant. It may be described as "making a particular many-to-few mapping of the the text" (Stone, et al., 1966, p. 7). The process has been described by Fisher (1968, p. 2):

> In particular, we attempt to reduce the complexity of a text to a limited set of potentially understandable events. To do this, we consider that a given meaning or theoretically defined quality (the numenon) may occur more than once, and that it may assume different forms of appearance (phenomena).

In the terminology of content analysis, a numenon is called a "tag," the grouping of related phenomena is called a "category," and the grouping of categories relevant to a given study is called a "dictionary." The task of the content analyst is to code the manifold elements of a communication into instances of category membership. The range of possible categories is practically infinite, but the ultimate criterion for selection is quite simple. Those categories should be utilized which allow the investigator to answer the questions he is asking.

As Zieky (1968, p. 18) noted: "The exact manner in which elements of the text are to be reduced to category scores has become a major source of contention among researchers active in the field." Proponents of a qualitative coding procedure (George, 1959; Kracauer, 1952) hold that an emphasis on numerical analysis restricts the range of the problems amenable to content analysis, and forces the researcher to ignore what might prove to be the most important variable in his source. In Kracauer's opinion, "many quantitative investigations . . . mark the spot where a misplaced desire for objectivity has failed to reveal the inner dynamics of atomized content" (1952, p. 642).

The advocates of quantitative research (Lasswell, et al., 1952; Cartwright, 1953; Stone, et al., 1966), however, present a very strong case for their position. Their arguments stress the precision, replicability and generalizability of their results. The advantages of the quantitative procedures were promulgated by Budd (1963, p. 25).

> Quantification in content analysis, as in other research, leads eventually to summarizing procedures resulting in some sacrifice of detail.... What is gained, of course, is more valuable. For the analyst in reality has lost nothing.... He has traded some unmanageable data for manageable information; he has exchanged his 'lost' data for efficiency and scientific rigor.

Even though the debate concerning the relative merits of quantitative and qualitative coding procedures has occupied a good deal of the time of those concerned with content analysis, the dichotomy is actually spurious (Holstoi, 1966). As Pool (1959, p. 192) remarked: "It should not be assumed that qualitative methods are insightful, and quantitative ones merely mechanical methods for checking hypotheses. The relationship is a circular one; each provides new insights on which the other can feed."

The subject of all content analysis is some mode of communication. The object of the research, however, may be information concerning either: (1) the communication itself; (2) the source of the communication; or (3) the effects of the communication.

Studies of the first type have been used most frequently in the history of the research methodology. They require the least amount of inference since the researcher is concerned mainly with the attributes of the message itself; but it should be emphasized that such studies are not limited to mere descriptive statements (Zieky, 1968). Communication may be compared to some relevant standard; communications from the same source may be compared across time or across situations of prediction; or communications from two or more sources might be investigated in parallel (Holstoi, 1966).

Studies which attempt to ascertain "Lawful relations between events or messages and processes transpiring in the individuals who produce . . . them" are clearly of the second type (Osgood, 1959, p. 36). Inferences are based on the assumption that certain attributes of the originator of a communication are reflected in variables existing within that communication. In effect, this also is the basis for the scoring of tests, for from the answers that are given to questions appearing on the test, inferences are made about the presence or absence of traits in the examinee. Moreover, increases in the amount of verbalization that constitute responses to the test are accompanied by increases in the number of inferences which must be drawn by the test scorers. It is the second type of content analysis, then, around which the present research centers.

The third type of study, though theoretically as important as the first two, has been relatively rare. It seems probable that direct behavioral observation of the receiver of a message has been a more efficient way of discovering the effects of that message than content analysis has been (Zieky, 1968).

No matter which of the above paradigms is chosen by the researcher, once category construction is complete he is faced with the task of transforming his data to instances of category membership. This process, known as coding, tends to be extremely slow, expensive and inefficient. The difficulties involved in the coding process led Berelson (1952, p. 198) to warn that unless there were particularly good reasons for using content analysis, "it is not worth going through the rigor of the procedure, especially when it is so arduous and so costly of effort."

The reduction of a communication for the purposes of content analysis, much like the scoring of many extant verbal-oriented tests, is tedious, repetitive, painstaking work which requires a large number of discreet comparisons, followed by decisions based upon those comparisons. The very nature of the task which makes it onerous for human judges makes it perfectly open to automation. Two sets of computer programs now exist which will perform the coding process quickly and accurately: The General Inquirer (Stone, et al., 1966) and SCORTXT (Fisher, 1968).

The advantages of the computer for content analysis are manifold. Once text is put into machine-readable form, any number of analyses might be attempted at a small increase in cost. The computer functions as a perfectly reliable judge which does not suffer from fatigue or lapses of attention. But perhaps an even more beneficial aspect of automation lies in the precise specification of both category members and coding procedures required of the researcher. In other words, the content analyst must know exactly what he is doing before he is able to instruct a computer to perform the same task. Savings in time are, of course, considerable; and the researcher is freed for attention to intellectual rather than mechanical problems.

It must be stressed that the introduction of automated procedures to perform the coding in no way releases the content analyst from the responsibility of carefully performing the other aspects of the research. The computer will count whatever it has been told to count; and it is dangerous to assume that lists of numbers are impressive, "scientific," or correct, merely because they were generated by a computer rather than a human judge. Both SCORTXT and General Inquirer replace one step of a content analytic study: that of reducing the communication to instances of category membership. If the data has been poorly sampled, if the categories have not been well constructed, or if the statistical analysis is inadequate or erroneous, the use of a computer to perform the coding will be a waste of time, money, and effort. The computer in content analysis is a tool used by the researcher, not a means of replacing the fallible human judgment upon which the quality of the study depends.

An understanding of the wide scope of the applications of content analysis may best be gained through an examination of selected studies. During approximately the first third of the twentieth century, content analysis was centered on journalism. Using such variables as number of column inches devoted to a topic, headline size and location, Speed (1893), Garth (1916), and Willey (1926), for example, were able to empirically substantiate their hypotheses concerning changes in focus, emphasis, and purpose of the newspapers in which they were interested.

The utility of content analysis outside of mass media research was demonstrated in the studies of enemy propaganda and political speeches which accompanied World War II. Under pressure to make valid predictions of enemy actions, researchers made advances in such methodological aspects as sampling theory, category, validity, coder reliability and syntactic analysis (Lasswell, et al., 1952).

The continuing interest in the problems noted above has been documented by Barcus (1959) in his survey of the literature. He found that the number of studies done using content analysis has approximately doubled each decade since 1930. It seems probable that this rate will increase as the use of computerized scoring procedures becomes more widely accepted. One measure of the wide use which content analysis has achieved in recent years is the diversity of the studies reported in The General Inquirer (Stone, et al., 1966). The relationship between personality, culture, and themes of folktales are analyzed, as are the distinguishing features of Presidential Nomination Acceptance Speeches, and the differences between the language of normal and psychotic subjects. There are also investigations of suicide notes, product images, therapy interviews, and Huckleberry Finn.

Although research has not been as expansive as one might expect, studies reporting the application of automated content analysis to the scoring of verbal responses to elected tests have also been found in the literature. Two studies utilizing the SCORTXT program for the scoring of teacher-constructed tests have been reported (McManus, 1968; Marcotte, 1969). An analysis intended to duplicate the hand-scoring system of content analysis for responses to the Thematic Apperception Test (McClelland, 1953) also has been attempted, using The General Inquirer (Stone, et al., 1966). The researchers, who in this study chose to score responses for the occurrence of one theme, Need Achievement, have commented (Stone, et al., 1966, pp. 192-193):

> By so limiting our interest we can attemtp to construct a system that would answer the question, Does the sentence, paragraph, document, and so forth, contain X theme or does it not? This, of course, is not an artificial problem. In fact, a number of hand-scoring systems have been developed to answer this type of question. With this in mind, we decided to approach the task just set forth by attempting to duplicate a hand-scoring system on the computer. Therefore, our goal was to construct rules enabling the computer to make decisions that are reliably similar to the decisions of a skilled judge. In making such an attempt, we have had in mind the following question: At our present level of technology, what are the possibilities of computers making interpretive judgements? As we come nearer to an answer to this question we also hope to illuminate more sharply some of the difficulties inherent in translating rules governing human strategies of decision into rules governing automatic strategies of decision.

No attempt to apply similar strategies to the scoring of the Torrance Tests of Creative Thinking (1966a) has been found in the literature.

Creativity

Since the last decade when Guilford (1950) called attention to the virtual neglect of the concept of creativity by American researchers, there has been an enormous expansion of interest and research in the nature of this higher mental process. Due to the plethora of relevant research which has followed Guilford's article, the review of the literature presented here makes no pretense at being complete in its coverage. ~~However~~, a sampling of the research most pertinent to this study will be presented.

Many possible (and somewhat different) definitions for creativity have been offered. For example, Torrance (1966c) has defined creativity as:

> A process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on; identifying the difficulty; searching for solutions, making guesses, or formulating hypotheses about the deficiencies; testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results.

Although this definition allows the types of abilities, mental functions, and personality characteristics that are associated with creativity to be operationally defined, not all scholars agree with Torrance. Ausubel (1963), for example, objects on the grounds that creativity as a highly specialized and substantive capacity cannot be distinguished from general intellectual abilities, personality variables, and problem-solving traits. On the other hand Ghiselin (1955) proposes simply that the measure of a creative product be the extent to which it restructures our universe of understanding. The working definition used by Stein (1955) is that a process is creative when it results in a novel work that is accepted as tenable or useful or satisfying by a group at some point in time. Others strongly feel that creativity measurement should only be employed when referring to such specialized fields as art, music, and writing (Kreuter and Kreuter, 1964; Mueller, 1964). It is clear that no single definition of creativity satisfying all the workers in the field has yet emerged. However, it is agreed that a psychological trait which generally can be referred to as creativity does exist, and that it exists in everyone to some extent (Lowenfeld, 1959; Hallman, 1963).

Although creativity does exist, the measurement of this higher mental process has not been an easy task, for a number of reasons. However, a variety of methods of assessing creativity has been attempted. Buel (1961), for example, used a number of personality instruments including the Kuder Preference Record (Kuder, 1953) to measure creativity. A study of the relationships between emotional stability, as measured by the Rorschach protocols, and creativity was made by Hammer (1961). An empirical study of the concept constancy construct of Asher's neo-field theory in which the Concept Constancy Test (1963) was used to assess creativity has also been reported (Jacobson and Asher, 1963). Studies in which the relationship between creativity and a multiplicity of variables including movement responses (Griffin, 1958), novelty of stimuli (Houston and Mednick, 1963), and "incidental stimuli" (Mendelsohn and Griswold, 1964) have also been reported. Despite these many attempts, an acceptable instrument for the measurement of creativity has not been isolated.

The work of Guilford and his associates has contributed much to the measurement of creativity. Using his theoretical structure of the human intellect as the research paradigm, Guilford (1967, p. 312) has found that problem solving and creative production are basically the same phenomena. He has chosen, therefore, to study creativity through problem solving, continually stressing the importance of divergent production tests, that is, tests that require examinees to produce their own answers rather than choose among alternatives.

The Torrance Tests of Creative Thinking (1966a) which in many respects may be regarded as a breakthrough in the area of creativity measurement, represents a rather sharp departure from the factor type tests developed by Guilford and his associates. Torrance and his colleagues have attempted to construct test activities that are "models of the creative process, each involving different kinds of thinking and each contributing something unique to the battery under development" (1966c, p. 9). However, the products that result from the administration of these tests are assessed in terms of Guilford's divergent thinking factors (fluency, flexibility, originality and elaboration). A further description of the Torrance Test of Creative Thinking will be presented in the following chapter, along with evidence for the reliability and validity of the test.

Summary

As previously mentioned, an exhaustive review of all of the literature dealing with both content analysis and creativity is clearly beyond the scope of this report. To provide some frame of reference, however, the most relevant research bearing on the present investigation was chosen for presentation.

Although a considerable amount of research in the possible uses of content analysis has been performed, the application of computerized content analytic techniques to the scoring of tests of creativity (in particular the Torrance Tests of Creative Thinking) has not been previously attempted. From the literature reviewed here, however, the use of the technique does appear to be a justifiable approach to the problem of analyzing the content of verbal responses elicited by creativity tests.

CHAPTER III

PROCEDURES

Introduction

This chapter will discuss the procedures which have been employed in this research. Included will be a description of the sample of students whose responses were analyzed; a discussion of the reliability and validity of the TTCT, as well as a general description of the instrument itself; a description of the techniques used in coding the data so that they would be recognizable by the computer; and a detailed description of the training of the human judges who would provide the criteria against which the automated scoring could be gauged, as well as a presentation of the pooled reliabilities of the judges for each of the various criteria. The steps taken in the construction of content analysis dictionaries will also be discussed, and a description of the computer program used to perform the automated content analyses will be presented. Finally, the scoring strategy which utilized the SCORTXT program will be described.

Description of the Sample

The sample used in this study consisted of pupils from 16 classes in grades four, five, six, and seven from six public school systems in central New York State. Classes were initially selected on the basis of similarity with respect to distribution of verbal intelligence scores, distribution of socio-economic class levels, reading levels, and sex (that is, similar proportions of boys and girls among classes at any grade level). On the basis of these criteria, four classes at each grade level were selected, yielding a sample of 375 students in 16 classes. From this large sample 153 students were selected for participation in the present study. Table 1 summarizes by grade level the mean, standard deviation, and range of scores on the Lorge Thorndike Verbal Intelligence Test (Lorge and Thorndike, 1962). The data on the sex of the participating pupils are summarized in Table 2. As can be seen in Table 2, the number of students participating differed from one grade level to the next. This was anticipated, however, since pupils were randomly selected from the total pool of data rather than for each grade level as such.

To allow for cross-validation processes the total sample of 153 subjects was randomly divided into two samples of 100 subjects and 53 subjects respectively (Mosier, 1951). The sample of size

TABLE 1

LORGE THORNDIKE VERBAL INTELLIGENCE TEST
SCORES BY GRADE LEVEL
FOR TOTAL SAMPLE[1]

Grade	N	Mean	SD	Range
4	19	110.06	18.97	76-137
5	33	110.50	16.82	95-143
6	50	106.98	12.43	73-134
7	51	112.19	13.64	88-138

TABLE 2

SUMMARY OF SEX DISTRIBUTION
BY GRADE LEVEL FOR TOTAL SAMPLE

Grade	Boys	Girls	Total
4	7	12	19
5	16	17	33
6	23	27	50
7	20	31	51

[1] In grades four through six, level three, form A, verbal: in grade seven, level four, form A, verbal.

100 served as the developmental sample for the computerized scoring procedure, while the cross-validation of these procedures was performed on the sample whose size was 53.

Instrumentation

Description of the Torrance Test

The measure of verbal creativity employed in this research was the Torrance Tests of Creative Thinking, Verbal Form A (Torrance, 1966a). This battery consists of seven parallel tasks of activities believed to bring into play somewhat different mental processes requiring the subject to think in divergent directions in terms of possibilities. The tasks or activities include: asking questions about a drawing; making guesses about the causes of a pictured event; producing ideas of improving a toy; thinking of unusual uses of a cardboard box; and thinking of the varied possible ramifications of an improbable event. The subtests are administered in a paper and pencil format in one setting requiring approximately 45 minutes of actual test-taking time.

With the exception of Activity 6, all seven Activities yield three scores for each response made by the subject. Fluency, the first score given, "is defined as the total number of relevant responses, relevancy being defined in terms of the requirements of the tasks as set forth in the Instructions" (Torrance, 1966b, p. 15). For example, Activity 1, the Ask Activity, the Fluency score is the number of relevant questions asked. Questions that can be answered merely by looking at the picture, however, are not considered relevant, therefore, are not counted. The fluency scores for the remaining activities are determined by counting the number of relevant responses. To determine the Flexibility score, the second score given each response, a number of categories have been constructed by Torrance and his associates. A category here simply means a classification or grouping of like responses; that is, responses dealing with the same subject matter. For example, Activity 1 in which questions are asked about a drawing, 22 characteristics of the drawing have been isolated. Examples of these categories are: a) the description of the figure; b) the physical action of the figure; c) the general costume worn by the figure; and d) the emotions of the figure. Each response or question that is classifiable within one of the 22 categories is awarded one point for Flexibility. All replications of the usage of any category are deleted, and the total Flexibility score for any activity is the sum of the points awarded to the individual responses, that is, the number of different categories used.

For the first five activities, for which the number of categories for each activity ranges from 21 to 24, the flexibility score is obtained by employing the aforementioned procedures. The flexibility score for Activity 7, "Just Suppose", is defined as a change or shift in attitude or focus. The first response is not scored. Each shift or change in attitude or focus receives one point. Once a shift is credited, duplications do not receive additional credit.

The third score given, Originality, is based on the frequency of the response in the population. The most frequent responses are assigned weights of zero and one, and a listing of these responses can be found in the scoring manual. If "infrequent responses show any creative strength and get away from the obvious, an Originality score of two should be assigned" (Torrance, 1966b, p. 20). The Originality score for each activity is the sum of the Originality scores for each of the individual responses. In Activity 6, Unusual Questions, originality weights for frequent responses are not provided. Responses are classified as either questions which require simple answers, questions which require complex answers, or divergent questions. Each question is then deemed either personal or factual, and corresponding weight are then assigned. As Torrance notes (1966b, p. 42);

> The kind of originality involved here may be somewhat different from the kind of originality involved in the other activities, since it is not based on statistical infrequency of response. Experience has shown, however, that there is a high positive correlation between statistical infrequency and the scores assigned using the criteria listed above, adopted from Burkhart's work (Burkhart and Bernheim, 1963).

To summarize the scoring procedures three scores are reported for each subject. The Fluency and Originality scores represent the sum of the corresponding scores attained in each of the seven activities. The Flexibility score is computed similarly, but it is based on six activities. Torrance cautions against combining the three dimensions of creativity scores to obtain a grand total, although several studies have based a criterion for creativity on such a total.

The Reliability of the Torrance Test

The test-retest reliability of the Torrance Tests of Creative Thinking (TTCT) has been shown to be quite high. In a test-retest situation using alternate forms of the TTCT, Torrance has reported reliability coefficients of .93 for Fluency, .84 for Flexibility, and .88 for Originality when the interval between testings was from one to two weeks. When the test interval was eight months, reliabilities of .79 for Fluency, .61 for Flexibility and .73 for Originality were found (Torrance, 1966c). It appears, then, that the size of the reliability coefficient is inversely related to the time between testing in the test-retest situation. It is important to note, however, that the creative abilities measured by the TTCT are susceptible to development through educational experiences. This would indeed influence the reliabilities in the manner previously stated. Goralski (1964) has reported reliability coefficients of .82, .78, and .59 for Fluency, Flexibility, and Originality when the interval was ten weeks. Mackler (1962) tested the same subjects three times with three different forms of the Ask-and-Guess test, each testing separated by a two-week interval. He obtained reliabilities of .82 (first and second testing), .89 (second and third testings), and .84 (first and third testings).

Both the inter- and intra-scorer reliabilities of the TTCT have also been shown to be very good. Torrance (1966c, p. 18) reports inter- and intra-scorer reliabilities to be "consistently above .90 and there have been only very small differences in means." In spite of these high reliability coefficients, Torrance points out the possibility of errors in scoring that exist when scorers fail to read carefully the scoring guide or to scan adequately the weights assigned to the Flexibility and Originality dimensions of Creativity (Torrance, 1966b). At least one reviewer (Hoepfner, 1967) has called attention to this problem. This shortcoming may be overcome, however, by the use of a computer programmed to score the TTCT, for, unlike humans, the computer can be counted on to be perfectly reliable.

The Validity of the Torrance Test

Evidence for the validity of the TTCT has been presented and discussed by Torrance (1966c). The evidence for the four principal aspects of validity-content validity, construct validity, concurrent validity, and predictive validity will be presented.

Making a case for the content validity of the TTCT Torrance states:

> A consistent and deliberate effort has been made to base the test stimuli, the test tasks, instructions, and scoring procedures on the best theory and research available. Analyses of the lives of indisputably eminent, creative people, research concerning the personalities of eminent creative people, the nature of performances of the human mind, and the like have been considered in making decisions regarding the selection of test tasks. A deliberate and consistent effort has also been made to keep the test tasks free of technical or subject matter content (1966c, p. 24).

Due to the complexity of creativity, Torrance does not believe that anyone can now specify the total range of test tasks necessary to completely assess or specify creative behavior. He does believe, however, that the test of tasks assembled in the battery do sample a wide range of the abilities of such a universe.

Evidence for the construct validity of the TTCT is presented by Torrance for children, high school youth, adults, and for studies of growth and learning. Since the present research utilized a sample composed of elementary school children, only those studies involving the use of the tests with elementary school children will be reviewed.

Weisberg and Springer (1961) compared the personality characteristics of highly creative and less creative fourth grade children, using the TTCT as the creativity criterion. The highly creative children were rated significantly higher than the less creative children on strength of self-image, ease of recall, humor, availability of Oedipal anxiety, and uneven ego development. These results, according to Torrance, reflect what might be called a creative acceptance of oneself and a greater self-awareness (1966c, p. 25). Rorschach protocols from the same research show that children of high creativity showed a tendency toward unconventional responses, unreal percepts, and fanciful and imaginative treatment of the blots. In addition, they also gave more human movement and color responses that the low group, signs regarded as indications of imaginativeness and creativeness among projective examiners.

In a study conducted by Torrance (1962) with highly creative children and less creative controls, three personality characteristics were found to differentiate the two groups. First, the highly creative youngsters produced significantly more "wild" or silly ideas. Second, their productions showed a high degree of Originality. Third, their productions were characterized by humor, playfulness, and relative relaxation. Fleming and Weintraub (1962) studied the relationship between rigidity and creativity as measured by the TTCT. Significant negative correlations were found between rigidity and Fluency, Flexibility, and Originality respectively. Yamamoto (1963) likewise reports significant correlations between creativity as measured by the Torrance tests and measures of Originality derived from evaluations of the imaginative stories of 20 fifth and 20 sixth graders.

In a study involving children in grades two through seven, Long and Henderson (1964) found that children high on creativity were able to withhold opinions under conditions of information inadequacy, withstand the uncertainty of an undecided state, and resist premature closure. Studies by Lieberman (1965), Long, Henderson, Ziller (1965), and Torrance (1963) present further evidence for the construct validity of the Torrance Tests of Creative Thinking.

Due to the lack of generally acceptable criteria, the concurrent validity of the TTCT has been difficult to assess. A number of criteria have been employed, however, and the results of these investigations do offer evidence worth noting.

Yamamoto (1964) analyzed the relationship between sociometric ratings and selected sub-tests of the TTCT. The sociometric questions used in this study were designed to measure dimensions of creative thinking ability. Statistically significant positive correlations were found, but this was due largely to the size of the sample (N for this study was 459).

Torrance (1962, 1963) and Yamamoto (1962) report studies with elementary school children in which the criterion for concurrent validation of the TTCT was teacher nominations. The results of these studies indicate that pupils nominated by their teachers as most Fluent, Flexible, and Original in their thinking can be differentiated by their scores on the TTCT from pupils nominated as lowest on these dimensions of creativity. The results reported by Nelson (1963), who had teachers use the Q-sort technique to rank students on creativity, agree with those reported by Torrance.

Another criterion measure employed in concurrent validity studies is educational achievement. Bish (1964) used the California Achievement Test scores as criteria in a study with fourth, fifth, and sixth grade youngsters. Correlations ranging from .36 to .42 were found between measures of Fluency, Flexibility, and Originality and measures of achievement. When IQ was partialled out these correlations increased. Cicirelli (1965) reports results similar to those found by Bish. Perry (1966) found statistically significant rank order correlations between the TTCT and subtests of the Stanford Achievement Test. The relationship between creativity and grades given in school was also investigated by Perry, and the correlation found (.10) was not statistically significant.

Since the Torrance Test of Creative Thinking is a relatively new instrument, only a limited amount of evidence for the predictive validity is available. Torrance reports, however, that a variety of long-range predictive validity studies are underway and that others have been planned for the near future (1966c, p. 54). In a study reported by Erickson (1966) creativity scores from the TTCT administered to high school seniors in 1959 were correlated with the scores from a checklist of creative activities administered in 1966. Forty-four of the sixty-six high school seniors who were tested in 1959 returned the checklist. Fluency and Flexibility scores correlated positively with the checklist criterion (.27 and .24 respectively) but Originality did not. The 44 subjects supplying checklist data were then divided into two equal groups on the basis of creativity test scores and tetrachoric correlations were computed for each item. Originality, Elaboration, and Total Creativity scores successfully predicted participation in a number of activities (writing a poem, story, song, or play; receiving a research grant; learning a new language, etc.) at acceptable statistical levels.

Despite the lack of rigerous evidence for predictive validity, the use of the TTCT in various forms of research does seem warranted as Torrance has claimed (1966c). Further, it does appear that the cautious use of the TTCT in a school setting is appropriate, since no better instrument for the assessment of creative potential is now available.

Data Collection and Coding Procedures

The data analyzed in this research consist of the responses of 153 elementary school children to the Torrance Test of Creative Thinking, Verbal Form A (Torrance, 1966a). The tests were administered in a group setting in accordance with the guidelines set forth in the test manual (Torrance, 1966c).

To perform the computerized content analyses of the data it was necessary to transcribe the responses into machine readable form. This was accomplished by keypunching the responses on standard IBM cards, one response to a card. Since no corrections in spelling, punctuation, grammar, etc., were made on the original copy, the keypunched data were an exact duplicate of the responses given in the test booklets. A sample of the keypunched data is shown in Figure 1. The first card listed in the Figure is a title or header card indicating that the responses which follow were given by the subject whose ID number was 0397. Following the title card are the responses given by the subject. For each response the ID number is punched in columns 1-4, the Activity number, which is "1" for the illustrated listing, is punched in column 5, and the response is given in columns 8-80. This subject made five responses to Activity 1, hence, the numbers in columns 6-7 range from 01 to 05. The card with the asterisks in columns 1-2 indicates the end of the responses for this subject. Note that for response five, the word "moment" is keypunched as "momet" since this was the spelling given in the test booklet. Also, although this response is stated in question form, no question mark was supplied by the subject. This, too, is reflected in the keypunching.

Human Test Rating Procedures

The Judges

To provide criterion measures against which the performance of the computer scoring of the tests could be gauged, it was necessary to obtain human ratings of the data. Four educational psychology students originally were selected and employed for this purpose. Due to changes in the availability of personnel two of the four scorers were replaced after the completion of the first three activities.

Procedures for Training Judges

To provide uniformity or orientation and to improve inter-scorer reliability, a number of procedures were utilized in the training of the judges. To give a greater appreciation for the concept of

0397

0397101Where Did this Boy Come from?

0397102What is his name?

0397103How did he get there?

0397104Why is he there?

0397105What is his means of transportation at the momet.

**

FIGURE 1

A SAMPLE OF MACHINE-READABLE RESPONSES TO THE TORRANCE TESTS OF CREATIVE THINKING creativity by becoming actively involved in the creative process, each judge was administered the Torrence Tests of Creative Thinking, Verbal Form A. Next, a series of seminars were conducted for the scorers during which the process of creativity and possible problems relating to the scoring procedures were discussed. The scorers were then provided with copies of Torrance's Guiding Creative Talent (1962) and were asked to read selected chapters. Copies of the Torrance Tests of Creative Thinking: Norms-Technical Manual (Torrance, 1966c) and the Torrance Tests of Creative Thinking: Directions Manual and Scoring Guide (Torrance, 1966b) were also provided. After the literature and manuals had been read, the judges were asked to score a sample set of responses listed in the Scoring Guide. The scorers then met as a group and discussed their rationale for assigning scores to each of the individual responses. Where differences of opinion existed between the judges and the Scoring Guide, the possible reasons for such differences were analyzed. As a final activity in the training process, a meeting was arranged between the scorers and Dr. E. Paul Torrance. During this meeting the scorers had the opportunity to raise any unresolved questions emanating from the practice scoring which they had performed.

Additional steps taken to improve reliability included: a) a discussion of the optimal amount of time for scoring in any one sitting; b) the provision of a "paste-up" of the scoring manual that enabled the scorers to view one activity or sub-test at a glance; and c) the scoring of the responses of all subjects to one activity before proceeding to the next acticity.

The actual scoring of the data was accomplished by the judges both individually and in group sessions in 1969.

Reliability of the Judges

The estimate of the reliability of the four judges taken as a group was determined for each of the activities for the total sample, the developmental, and the cross-validation sample respectively. For each activity the reliabilities for Fluency, Flexibility, and Originality were computed separately (as previously noted, there is no Flexibility score for activity six). The analysis of variance technique with judges acting as treatments (Winer, 1962, pp. 124-132) was used to obtain these estimates. The pooled reliabilities for each activity-dimension combination are reported in Tables 3, 4, and 5.

With the possible exception of the pooled reliability for the Originality dimension for Activities 3, 6, and 7, the reliabilities found for each set of four judges were higher than expected.

In an effort to determine the response tendencies of each of the scorers the means and standard deviations of the judges scores for the respective activity-dimension combinations were obtained. These may be found in Tables 6, 7, and 8. As can be seen, these means and

TABLE 3

RELIABILITY ESTIMATES FOR JUDGES USING

ANALYSIS OF VARIANCE TOTAL SAMPLE

Activity	Dimension	Between Subjects		Within Subjects		Between Judges		Residual		Reli-ability
		df	ms	df	ms	df	ms	df	ms	
1	Fluency	152	58.04	459	.53	3	3.54	456	.51	.99
1	Flexibility	152	24.29	459	.55	3	1.14	456	.55	.98
1	Originality	152	40.84	459	7.77	3	472.39	456	4.72	.81
2	Fluency	152	30.43	459	1.26	3	10.78	456	1.20	.95
2	Flexibility	152	10.64	459	.71	3	6.67	456	.67	.93
2	Originality	152	41.86	459	8.50	3	471.37	456	5.46	.80
3	Fluency	152	42.98	459	2.80	3	25.44	456	2.65	.93
3	Flexibility	152	11.17	459	.86	3	4.10	456	.84	.92
3	Originality	152	50.84	459	17.29	3	1292.85	456	8.90	.66
4	Fluency	152	170.86	459	3.89	3	70.33	456	3.45	.98
4	Flexibility	152	25.42	459	1.13	3	12.53	456	1.06	.96
4	Originality	152	71.06	459	16.77	3	1093.62	456	9.68	.76
5	Fluency	152	283.72	459	14.87	3	175.83	456	13.81	.95
5	Flexibility	152	43.46	459	1.82	3	5.25	456	1.80	.96
5	Originality	152	118.91	459	42.45	3	2800.59	456	24.31	.64
6	Fluency	152	69.09	459	.63	3	6.94	456	.59	.99
6	Originality	152	39.66	459	15.84	3	346.71	456	13.66	.60
7	Fluency	152	31.86	459	1.99	3	6.61	456	1.96	.94
7	Flexibility	152	10.00	459	1.18	3	.78	456	1.19	.88
7	Originality	152	4.13	459	1.23	3	25.34	456	1.06	.71

TABLE 4

RELIABILITY ESTIMATES FOR JUDGES USING

ANALYSIS OF VARIANCE DEVELOPMENTAL SAMPLE

Activity	Dimension	Between Subjects		Within Subjects		Between Judges		Residual		Reli-ability
		df	ms	df	ms	df	ms	df	ms	
1	Fluency	99	64.28	300	.46	3	2.08	297	.44	.99
1	Flexibility	99	23.80	300	.56	3	1.08	297	.56	.98
1	Originality	99	45.93	300	8.42	3	317.01	297	5.31	.81
2	Fluency	99	30.07	300	1.10	3	10.89	297	1.01	.96
2	Flexibility	99	10.97	300	.74	3	5.49	297	.69	.93
2	Originality	99	43.31	300	7.02	3	262.22	297	4.45	.84
3	Fluency	99	43.72	300	2.25	3	14.48	297	2.13	.94
3	Flexibility	99	11.58	300	.77	3	1.72	297	.76	.93
3	Originality	99	59.88	300	16.27	3	807.42	297	8.28	.73
4	Fluency	99	185.09	300	4.30	3	42.34	297	3.92	.98
4	Flexibility	99	23.70	300	1.16	3	7.69	297	1.10	.95
4	Originality	99	78.10	300	19.37	3	817.67	297	11.31	.75
5	Fluency	99	269.18	300	17.75	3	142.29	297	16.49	.93
5	Flexibility	99	43.00	300	1.94	3	7.58	297	1.89	.96
5	Originality	99	97.93	300	40.00	3	1763.69	297	22.59	.60
6	Fluency	99	75.00	300	.75	3	5.38	297	.71	.99
6	Originality	99	49.75	300	18.36	3	219.30	297	16.33	.63
7	Fluency	99	37.45	300	2.35	3	6.68	297	2.31	.94
7	Flexibility	99	4.43	300	1.43	3	21.09	297	1.23	.68

TABLE 5

RELIABILITY ESTIMATES FOR JUDGES USING

ANALYSIS OF VARIANCE CROSS-VALIDATION SAMPLE

Activity	Dimension	Between Subjects		Within Subjects		Between Judges		Residual		Reli-ability
		df	ms	df	ms	df	ms	df	ms	
1	Fluency	52	47.29	159	.66	3	2.01	156	.63	.99
1	Flexibility	52	25.61	159	.55	3	1.32	156	.51	.98
1	Originality	52	31.85	159	6.55	3	158.96	156	3.62	.79
2	Fluency	52	31.49	159	1.56	3	1.98	156	1.55	.95
2	Flexibility	52	10.21	159	.67	3	2.06	156	.64	.93
2	Originality	52	39.44	159	11.29	3	221.22	156	7.25	.71
3	Fluency	52	41.92	159	3.83	3	13.35	156	3.65	.91
3	Flexibility	52	10.23	159	1.02	3	6.39	156	.91	.90
3	Originality	52	34.60	159	16.60	3	518.46	156	9.63	.52
4	Fluency	52	143.02	159	3.11	3	28.49	156	2.62	.98
4	Flexibility	52	28.83	159	1.08	3	5.51	156	.99	.96
4	Originality	52	56.88	159	11.85	3	286.14	156	6.57	.79
5	Fluency	52	311.58	159	9.44	3	49.15	156	8.68	.97
5	Flexibility	52	45.11	159	1.59	3	2.73	156	1.57	.97
5	Originality	52	161.05	159	47.09	3	1076.20	156	27.30	.71
6	Fluency	52	55.05	159	.39	3	1.69	156	.36	.99
6	Originality	52	19.08	159	11.08	3	136.28	156	8.68	.42
7	Fluency	52	18.87	159	1.30	3	1.98	156	1.29	.93
7	Flexibility	52	8.80	159	.83	3	.74	156	.84	.91
7	Originality	52	3.28	159	.82	3	5.30	156	.73	.75

TABLE 6

MEANS AND STANDARD DEVIATIONS FOR JUDGES

TOTAL SAMPLE

Activity	Dimension	Judge 1		Judge 2		Judge 3		Judge 4	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1	Fluency	5.83	15.70	6.02	14.32	5.83	15.14	5.65	14.41
1	Flexibility	4.16	6.46	4.27	5.96	4.26	6.86	4.09	6.66
1	Originality	1.61	2.50	3.33	9.66	5.87	27.26	3.96	15.59
2	Fluency	5.57	7.87	5.48	8.57	5.03	9.01	5.61	8.61
2	Flexibility	3.84	3.02	3.93	3.49	3.45	3.22	3.80	2.95
2	Originality	2.40	5.33	4.07	14.41	6.65	21.70	4.64	16.81
3	Fluency	5.42	12.86	6.15	11.51	5.96	12.16	6.37	14.42
3	Flexibility	3.45	4.18	3.63	3.01	3.61	3.27	3.85	3.23
3	Originality	1.58	3.06	3.71	8.08	7.60	29.77	7.28	36.66
4	Fluency	11.80	6.99	11.54	6.68	12.30	7.15	10.68	6.04
4	Flexibility	5.84	2.67	5.75	2.72	5.91	2.73	5.28	2.56
4	Originality	2.37	2.38	8.59	6.31	5.08	4.60	6.99	5.78
5	Fluency	13.90	8.45	16.01	9.82	13.74	8.85	13.97	8.89
5	Flexibility	7.39	3.54	7.42	3.55	7.04	3.31	7.15	3.57
5	Originality	3.84	3.56	11.13	8.45	3.23	3.59	10.72	9.74
6	Fluency	5.33	4.19	5.63	4.19	5.84	4.22	5.64	4.23
6	Originality	1.32	2.64	4.48	5.94	2.87	4.56	1.33	4.20
7	Fluency	4.75	3.23	4.37	3.24	4.28	2.84	4.39	2.95
7	Flexibility	1.99	1.74	2.03	1.99	1.94	1.90	1.86	1.73
7	Originality	.86	1.07	1.72	1.51	.95	1.36	.91	1.36

TABLE 7

MEANS AND STANDARD DEVIATIONS FOR JUDGES

DEVELOPMENTAL SAMPLE

Activity	Dimension	Judge 1		Judge 2		Judge 3		Judge 4	
		$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
1	Fluency	5.87	17.39	5.97	16.27	6.87	16.50	5.63	15.47
1	Flexibility	4.08	6.28	4.15	6.25	4.29	6.53	4.06	6.44
1	Originality	1.45	2.27	3.42	11.68	5.79	31.22	3.85	16.69
2	Fluency	5.42	8.13	5.43	7.95	4.85	8.92	5.61	8.12
2	Flexibility	3.80	3.21	3.91	3.50	3.40	3.25	3.87	3.10
2	Originality	2.28	5.84	4.12	13.74	6.22	20.13	4.56	16.96
3	Fluency	5.35	12.69	6.07	11.42	5.70	11.67	6.19	14.34
3	Flexibility	3.44	4.45	3.58	2.91	3.36	3.16	3.65	3.36
3	Originality	1.63	3.57	3.75	8.59	7.12	30.79	7.61	41.78
4	Fluency	12.17	7.30	11.98	6.93	12.72	7.46	11.15	6.30
4	Flexibility	5.91	2.62	5.94	2.60	6.02	2.68	5.41	2.48
4	Originality	2.47	2.35	9.12	6.69	5.31	5.01	7.37	6.05
5	Fluency	14.45	8.33	16.67	9.73	14.33	8.76	14.12	8.81
5	Flexibility	7.49	3.54	7.58	3.57	7.13	3.31	7.01	3.53
5	Originality	3.81	3.59	11.41	7.98	3.32	3.31	10.14	8.84
6	Fluency	5.73	4.39	6.06	4.36	6.29	4.40	6.09	4.42
6	Flexibility	1.69	3.05	4.63	6.35	3.40	5.00	1.52	4.91
7	Fluency	5.16	3.51	4.83	3.57	4.55	3.02	4.71	3.19
7	Flexibility	2.21	1.73	2.21	2.00	2.14	2.00	2.11	1.83
7	Originality	.92	1.04	1.92	1.63	1.10	1.50	1.02	1.46

TABLE 8

MEANS AND STANDARD DEVIATIONS FOR JUDGES

CROSS-VALIDATION SAMPLE

Activity	Dimension	Judge 1		Judge 2		Judge 3		Judge 4	
		$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
1	Fluency	5.76	12.80	6.11	10.87	5.76	12.84	5.68	12.68
1	Flexibility	4.30	6.91	4.49	5.44	4.19	7.62	4.13	7.19
1	Originality	1.93	2.84	3.15	5.94	6.02	20.21	4.17	13.72
2	Fluency	5.85	7.40	5.59	9.90	5.38	9.16	5.60	9.70
2	Flexibility	3.91	2.70	3.96	3.54	3.55	3.21	3.66	2.69
2	Originality	2.62	4.40	3.98	15.94	7.47	24.06	4.79	16.82
3	Fluency	5.55	13.41	6.30	11.87	6.45	12.95	6.71	14.67
3	Flexibility	3.47	3.75	3.71	3.25	4.09	3.16	4.23	2.83
3	Originality	1.47	2.14	3.64	7.23	8.51	27.10	6.66	27.04
4	Fluency	11.09	6.38	10.70	6.16	11.51	6.52	9.79	5.45
4	Flexibility	5.70	2.78	5.40	2.93	5.70	2.85	5.02	2.71
4	Originality	2.19	2.44	7.59	5.45	4.64	3.69	6.26	5.23
5	Fluency	12.87	8.65	14.76	9.96	12.62	8.99	13.70	9.10
5	Flexibility	7.21	3.56	7.11	3.53	6.87	3.35	7.42	3.67
5	Originality	3.89	3.56	10.60	9.32	3.06	4.09	11.81	11.25
6	Fluency	4.57	3.69	4.83	3.78	5.00	3.77	4.79	3.74
6	Originality	.62	1.36	4.19	5.11	1.87	3.41	.98	2.34
7	Fluency	3.98	2.48	3.51	2.30	3.77	2.42	3.77	2.33
7	Flexibility	1.57	1.70	1.68	1.93	1.57	1.64	1.40	1.42
7	Originality	.76	1.12	1.34	1.39	.68	1.00	.70	1.14

standard deviations are highly similar for the Fluency and Flexibility dimensions. Some substantial differences may be noted for the Originality scores, which is not surprising since the assessment of Originality is the most difficult task a judge must perform in scoring the TTCT. This suggests that a weighted composite of the judges' scores might yield a more reliable estimate of Originality. Results reported in a derivative study by Greene (1970), where factor analytic techniques were employed to derive differential weightings for judges, do not indicate great increases in the reliability of those scores. It appears, then, that the use of a simple composite of judges scores as the criteria for this research is warranted.

The Tags

The isolation of a set of concepts to be used as tags or category headings is a process that is of major importance to the outcome of content analysis research. Zieky (1968, p. 38) observed:

> If the most rigorous of validating procedures is used to confirm category membership, if the coding procedure is perfectly reliable, and if good statistical techniques are used to analyze the data, the research will nevertheless be of little use unless those concepts represented by the categories are relevant and theoretically meaningful.

Usually, the researcher at this point is faced with two choices: he may either isolate his own set of concepts, or utilize those represented by a pre-existing content analysis dictionary. In the present research, however, tags had been isolated, but the dictionaries associated with these concepts had not been constructed (Archambault, 1969).

As previously mentioned, Torrance and his associates have isolated sets of concepts to be used in the scoring of each activity of the TTCT. That the concepts are relevant and theoretically meaningful has been argued repeatedly by Torrance in the discussions of the various validity studies of the tests. This ideal situation indicated not only that these concepts would be used in the present research, but also that the techniques to be employed were both appropriate and justified.

For Activity 1, 22 concepts have been used in the assignment of Flexibility scores to the subjects' responses. A listing of these concepts is given in Figure 2. Since the meaning attached to each concept is explained in detail in the Scoring Guide, only the category headings have been given here. The 21 tags of Activity 2 and Activity 3 are the same as those listed for Activity 1 with the following exceptions: Tag 13 of Activity 1, Personal Possession or Past History of Figure, is deleted; Tag 22 of Activity 1, the Whole Picture, is deleted; Tag 17 of Activity 1, Reflective Surface and Reflection

1. Characters
2. Costume
3. Description of the Figure
4. Emotions
5. Ethnic Factors
6. Family
7. Ground Surface and Characteristics of Objects on it
8. Hat
9. Location
10. Magic
11. Occupation
12. Pants
13. Personal Possessions or Past History of Figure
14. Physical Action Related to Reflective Surface
15. Physical Action Unrelated to Water
16. Physical Characteristics of Objects or Situation
17. Reflective Surface and Reflection Itself
18. Shirt
19. Shoes
20. Time
21. Underwater
22. Whole Picture

FIGURE 2
LISTING OF THE CATEGORY HEADINGS USED IN
SCORING ACTIVITIES 1, 2, and 3, OF THE TTCT

NOTE: For Activities 2 and 3 tags 13 and 22 were deleted. Tag 17 was subdivided into two tags for Activities 2 and 3.

Itself, becomes two tags in Activities 2 and 3, one tag for the Reflective Surface and the other tag for the Reflection Itself.

The specific categories for Activities 4 and 5 are listed in Figure 3 and Figure 4. No flexibility category exists for Activity 6.

For Activity 7 isolated concepts were not provided and a shift in attitude was used to denote flexibility. Rather than approach this shift parameter directly, 39 categories were isolated (Greene, 1970) to function as an estimator of the shift parameter. These categories were derived by considering the originality weight list provided by Torrance for this activity and by examining the actual subject responses in the developmental sample; however, no specific name was assigned to each category, and they were merely numbered. The construction of the categories for each of these tags will be discussed in the next section.

In addition to the tags for Flexibility, it was necessary to designate concepts for the scoring of Originality. Since responses to the first three Activities of the TTCT are assigned Originality weights of either 0, 1, or 2, a tag for each weight was isolated. Actually, since the categories constructed for the Originality tags are combinations of entries from the Flexibility dictionaries, it would have been possible to weight the entries in the Flexibility categories with the appropriate Originality loadings. This procedure, however, would have necessitated complex modifications in the SCORTXT program. For this reason the isolation of tags for Originality seemed more appropriate.

The categories for Originality for Activities, 4, 5, and 7 were implicitly defined by Torrance. As will be explained in the next section, the Originality 0-weight has no value in the computerized scoring process, but it is automatically generated from the Flexibility categories. The process of assessing Originality in Activity 6, "Unusual Questions", is unique. The technique for categorizing the questions is presented again in Figure 5 with the corresponding point values. Rather than considering six categories, one for each factor or type of question, only three categories were generated. These categories were based on the point values 1, 2 and 4. Even though the Original six categories have been collapsed to three, no useable information has been sacrificed.

Category Contruction

When the tags to be used in the content analysis had been decided upon, the next step in the construction of the dictionaries was the assignment of words and phrase to the particular categories. The major theoretical basis for the word assignment was the simple and widely accepted principle of <u>synonymity</u>. The operational definition upon which the category construction was based has been given

1. Adaptation
2. Addition
3. Change Color
4. Change Shape
5. Combination
6. Division
7. Magnification
8. Minification
9. Motion
10. Multiplication
11. Position
12. Quality of Material
13. Rearrangement
14. Reversal
15. Ear Appeal
16. Touch Appeal
17. Eye Appeal
18. Smell Appeal
19. Substitution
20. Subtraction
21. Humanization

FIGURE 3

FLEXIBILITY CATEGORY HEADINGS FOR ACTIVITY 4

1. Animal Shelter
2. Animal Uses Other than Shelter
3. Art Uses
4. Buildings
5. Construction Uses
6. Carrier
7. Container
8. Costume
9. Cover
10. Destruction
11. Education
12. Furniture
13. Games
14. Growing
15. Household Appliances and Other Items
16. Protection
17. Scientific Uses and Equipment
18. Storage
19. Tools
20. Toy Furniture or Household Appliances
21. Toys
22. Transportation (Air)
23. Transportation (Surface)
24. Weapons

FIGURE 4

FLEXIBILITY CATEGORY HEADINGS FOR ACTIVITY 5

Type of Question	Personal	Factual
Simple Answer	1 point	0 points
Complex Answer	2 points	0 points
Divergent	4 points	4 points

FIGURE 5

ORIGINALITY CATEGORY HEADINGS AND WEIGHTS
FOR ACTIVITY 6

by Zieky: "Those words which are synonyms of a given word are listed under the main entry for that given word in a standard dictionary of English synonyms or a standard thesaurus" (1968, p. 42).

Sedelow, Sedelow, and Ruggles (1964, p. 220) established a precedent for the use of such a procedure:

> Because conventional thesauri are organized in terms of putative semantic relationships, we have chosen to use the thesaurus form as the basis for the VIA program. We take it, further, that semantic similarity is perceived in part in terms of word roots, i.e., words with the same root are likely to have meanings which have some connection with each other. Our VIA thesaurus, therefore, is constructed on the basis of (a) identical root, (b) synonymity and antonymity...

The procedure followed in the construction of categories, bases on the procedures listed by Zieky (1968, pp. 43-46), will be discussed in two sections. This is necessary since different operations were followed in the construction of the Flexibility and Originality categories. The following are the operations for the construction of the Flexibility categories:

(1) Each conceptual heading or tag was found in Roget's International Thesaurus (1962). For example, a search of the tag "magic" produced words such as "witchcraft," "spell," "sorcery," "charm," etc.

(2) The word list was copied on IBM cards, one word to a card, with the exclusion of those words marked as dialectical, vulgar, colloquial, jocular, or slang.

(3) The same procedure was followed with word lists that appeared under headings cross-referenced by the original main entry.

(4) The synonyms of each word were then found in Soule's Dictionary of English Synonyms (1966), and entered on IBM cards.

(5) The synonyms of the first-order synonyms were similarly retrieved to form a list of second-order synonyms. It was determined that retrieval of third-order synonyms was not profitable in terms of the ratio of additions to the category to the number of words retrieved.

(6) Since it is desirable for the categories to be mutually exclusive as has been stressed by Budd (1966) and Pools (1959), phrases were used to disambiguate the meaning of words which might appear in more than one category. For example, the word "play" is appropriate for both category 11 and category 13. For category 11, the phrase "play a part" was the meaning desired while "play in water" was the meaning intended for category 13. When the phrases

are employed instead of the single words the difference in meaning is obvious to both the reader and the computer.

(7) After the list derived from the thesaurus and synonym dictionary was completed and a check for errors was made, morphophonemic variants of root words were added to the synonym list. For example, the words "jumped" and "jumping" were added to the list which had formerly contained the word "jump."

(8) Throughout the category construction process listings of the responses of the developmental sample were consulted to determine if pertinent entries had been omitted. The precedent for this approach was established by Stone, et al. (1966, pp. 147-148):

> Thus the language characteristics of the target population of the study affect the dictionary by establishing which theoretical distinctions will be meaningful in practice and by affecting the assignment of entry words to particular tags on the basis of their usage. We increasingly use samples of the text to be analyzed in creating lists of entry words, determining the most appropriate definitions (on the basis of the usage exhibited in these samples from documents of this particular language community) and checking whether the levels of abstration of our tag categories are feasible. A dictionary is the product that emerges from the dual demands of theory on the one hand and concrete data on the other.

(9) The resulting list of words and phrases formed the category which was given another check for errors. These procedures were generalized to the remaining activities of the TTCT (with the exception of Activity 6). The categories were then keypunched on IBM cards in the format specified for the operation of the SCORTXT program. A computer listing of the dictionary was then attained and the keypunching verified.

The Originality categories (with the exception of Activity 6) were constructed from the Flexibility categories that had already been generated. Again, the criterion for inclusion in each category was synonymity. The steps included in the construction of these categories were:

(1) For each Flexibility category the weights given to the frequent responses listed in the Scoring Guide were analyzed. It will be recalled that frequent responses are given Originality weights of zero and one.

(2) For each response given an Originality weight of zero in the manual, the key word or phrase that would enable the scorer to classify this response in the particular Flexibility category was

extracted. This word or phrase would then be entered in the Originality category of zero weights. For example, for Activity 1 in Flexibility category 20 the response, "How old is he?", is given a zero Originality weight. Since Flexibility category 20 deals with time, the key word in this response is "old." The word "old," which is in the Flexibility dictionary, would, therefore, be entered in the zero Originality category.

(3) For each word or phrase entered in the zero Originality dictionary by "2" above, all synonyms of the word or phrase occurring in the Flexibility dictionary were extracted and included in the zero Originality category. As an example, the word "age" is a synonym of "old" and is included in the Flexibility dictionary. The word "age" should, therefore, be assigned an Originality weight of zero.

(4) The procedures of steps 2 and 3 above for responses given an Originality weight of one in the Scoring Manual were followed. These entries will constitute the Originality 1 category.

(5) Any entries of the Flexibility dictionary which have not been entered in either the zero or one Originality categories were entered in the Originality 2 category. This was justified under the guidelines set forth by Torrance who stated" "A judgment has to be made concerning the obviousness of a response when it is not included in the lists accompanied by Originality weights and Flexibility categories. Most responses not included in these lists should be given maximum Originality weights." (1966c, p. 20)

The basic structure for the three categories of the Originality dictionary of Activity 6 was generated by identifying the key phrases in the examples provided by Torrance. His definition of each type of question was also considered. Although the scoring procedure involved here is complex, a relatively simple device for the generation of Originality categories emerged. The appropriateness of this scheme was justified when the dictionary generated by it proved to be a most useful predictor.

Computerized Scoring Procedures Utilizing the SCORTXT Program

When all of the steps in the preparation of the data and the building of dictionaries had been taken, the next procedure to be completed was the automated reduction of subjects' responses to instances of category membership. That is, through the use of a computer programmed to perform the task, the verbal input was reduced to a series of numbers indicating scores for the various categories. Concurrent with this reductive process, average word length, average sentence length, number of periods, number of question marks and other acturaial measures were calculated. This entire procedure was performed by Fisher's SCORTXT program (1968).

The SCORTXT system consists of a main program and nine subroutines, all written PL/1. Although the program currently runs under the IBM 360 OS system, there is no machine dependence built into the program. The program itself has four sections. The first determines the run options for any particular analysis. These include the printing of text in string or array form, the removal or retention of punctuation, the choice of card margins to be scanned, the maximum storage size for the input text and the dictionary, and the printing of the item analyses for each category in the dictionary. In addition, punctuation counts and word length statistics can also be calculated by the program

The second section is devoted to the construction of a dictionary file. As mentioned, both words and phrases may be included in the dictionary. Dictionary entries are sorted internally into alphabetical sequence, so the categories may be entered in any order. The third section creates a text file of the data to be processed, which in this case are the students' responses to the TTCT. The actual scoring is done in the fourth section of the program by the use of a binary search algorithm. To the writers' knowledge, this is the only phrase lookup algorithm employing the binary search technique throughout. After the text has been processed, the category counts and the counts for the various word length statistics are punched out on IBM cards. Printed output is determined by the options that have been selected for any particular run. It should be noted that SCORTXT was devised to be generally applicable to a wide range of natural language analysis problems. Any of the subroutines may be used independently; any number of texts may be processed with any number of dictionaries; phrases as well as single words may be included in the dictionary, and those phrases need not be fully defined. For example, if phrases such as "playing near the water," "playing in the water," "playing with the water," are members of a dictionary; they could all be coded by the single entry "playing x the water": where x is defined as the 0-8-2 punch.

To maximize the prediction of each subjects' scores for each activity of the TTCT the step-wise multiple regression technique was employed. Since the scores for Fluency, Flexibility, and Originality are predicted differently, they will be discussed individually in this section. However, each of the approaches to be described is based on the rationale given by Torrance in the <u>Directions Manual and Scoring Guide</u> (1966b).

As mentioned previously, the Fluency score for each Activity is defined as the sum of the Fluency scores for the individual responses. Since no consideration is given at this point to the category in which the response falls, the punctuation and word length statistics calculated by SCORTXT were the only variables used in these predictions. Those SCORTXT variables considered were chosen by the following rule: each variable with a mean and standard deviation greater than one, i.e., those variables for which instances were actually found in the responses, were included in the prediction equations. Twenty-four of the 59 statistics calculated by SCORTXT were isolated by this rule.

Figure 6 gives a listing of these variables. The results and prediction equations will be presented and discussed in the following chapters.

The Flexibility score for each activity is defined as the sum of the Flexibility scores for each response, when replications for any category have been deleted. The category scores from SCORTXT were, therefore, used in the prediction of Flexibility. In addition, since high correlations were found between some of the actuarial measures and the Flexibility criteria, these also were used in the prediction. Those actuarial variables considered are the same variables isolated previously for Fluency. The results and prediction equations for Flexibility will also be presented and discussed in the following chapters.

The equation for the prediction of Originality was based on both the Originality dictionary counts and the actuarial measures. In the choice of relevant actuarial variables for the prediction of Originality, the reasoning was the same as given for Flexibility. Again, the results and prediction equations will be presented and discussed in subsequent chapters.

As mentioned earlier in this section, the automated scoring technique was based primarily on the scoring paradigm developed by Torrance. The inclusion of actuarial measures as prediction variables did deviate from the model, but the use of these measures in the prediction of the various scores did seem justified in light of the correlations between the predictors and the criteria. The validities of the multiple regression procedures are evaluated by means of standard cross-validation techniques.

Summary

Responses of 153 subjects to the Torrance Test of Creative Thinking, Verbal Form A were selected to serve as the data for the present research. To provide simulation criteria, the responses of each subject were rated by four trained human judges. High pooled reliabilities were found for the ratings given. The responses to the TTCT were keypunched on cards for subsequent evaluation by the computer, and dictionaries to be used in the automated scoring were constructed. The final strategies for the computerized scoring of the TTCT were also formulated.

Number of Question Marks

Number of Commas

Number of Periods

Number of Words of Length One

Number of Words of Length Two

Number of Words of Length Three

Number of Words of Length Four

Number of Words of Length Five

Number of Words of Length Six

Number of Words of Length Seven

Number of Words of Length Nine

Number of Words of Length Ten

Number of Words

Number of Sentences

Number of Paragraphs

Average Word Length

Average Sentence Length

Average Paragraph Length

Standard Deviation of Word Length

Standard Deviation of Paragraph Length

Third Moment of Word Length

Fourth Moment of Word Length

FIGURE 6

ACTUARIAL VARIABLES INCLUDED IN PREDICTION EQUATIONS FOR FLUENCY, FLEXIBILITY, AND ORIGINALITY

CHAPTER IV

RESULTS — ACTIVITIES 1, 2, AND 3

Since modificatons in the procedures for analyzing the results of this investigation were introduced into the analyses of the last four activities the findings will be reported in two separate chapters. Thus the more sophisticated statistical treatments for Activities 4 through 7 represent various approaches that tend to maximize prediction. The simpler and more utilitarian approach taken for Activities 1 through 3 are reported in the present chapter and are described separately for the reader who may wish to explore alternative strategies in similar types of research settings.

Establishing Prediction Equations

As indicated in the previous chapter, the step-wise multiple regression technique was employed to maximize the prediction of subjects' scores for each activity of the TTCT. Since nine scores were predicted for each individual, that is, a Fluency, Flexibility, and Originality score for each of the three activities, nine separate analyses were performed yielding nine different prediction equations.

The results of the step-wise multiple regression analysis for Activity 1, Fluency are presented in Table 9. Since Fluency is defined by Torrance as the number of appropriate responses given by the subject, it is not surprising that the variable "number of sentences," which is isomorphic to the number of responses, is the best predictor. It is surprising, however, that the multiple correlation coefficient is so high, .93, since no scheme for the determination of the appropriateness of the responses was incorporated in the scoring procedure. It is interesting to note the changes in the multiple-R coefficients when the number of predictors is increased in a step-wise manner. At Step 10, that is, when ten predictors are included, the multiple-R coefficient is .96 while at Step 24 a multiple-R coefficient of .97 is found, an increase of only .01. These data seem to indicate that the accuracy in prediction will not be significantly reduced by the elimination of some predictors. More will be said about this point in following chapters.

Tables 10 and 11 summarize the results of the multiple regression analyses for Fluency, Activities 2 and 3. Since the data are such that the number of paragraphs is equivalent to the number of sentences, the best predictor was again found to be the number of responses given. Although the sequence of variables entered into the regression equations is not the same for the three Fluency, criteria, the pattern of increase in the multiple-R coefficients is the same as previously noted.

TABLE 9

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 1 FLUENCY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	# sentences	.42281	15.33	.03	.926	1.52	593.66**
2	# question marks	.21047	.07	2.97**	.946	1.31	412.19**
3	words length 8	- 1.12423	.32	-3.54**	.950	1.27	297.70**
4	words length 4	.10336	.06	1.73	.952	1.25	231.21**
5	SD of sentence length	- .53224	.23	-2.31*	.953	1.24	187.77**
6	average sentence length	.31153	.21	1.52	.955	1.23	160.38**
7	# periods	- .11836	.09	-1.29	.956	1.22	139.97**
8	# commas	.72843	.30	2.40*	.957	1.20	124.73**
9	# paragraphs	.48173	15.33	.03	.958	1.20	112.38**
10	words length 7	- .30855	.13	-2.38*	.959	1.19	102.70**
11	SD of word length	1.86025	1.31	1.42	.961	1.18	95.36**
12	words length 1	- .21133	.20	-1.08	.961	1.18	88.33**
13	words length 3	.07585	.05	1.42	.962	1.18	81.81**
14	4th mom. of word length	- .42937	.43	- .99	.963	1.17	76.64**
15	# left parentheses	- 1.36839	1.88	- .73	.963	1.17	71.64**
16	average word length	- .58729	1.11	- .53	.963	1.17	67.19**
17	words length 14	3.09903	2.59	1.19	.964	1.17	63.71**
18	average paragraph length	64.53738	112.48	.57	.965	1.17	60.07**
19	SD of paragraph length	-27.50931	84.24	- .33	.965	1.18	56.32**
20	# ampersands	.69172	2.25	.31	.965	1.18	52.90**
21	3rd mom. of word length	.51919	1.62	.32	.965	1.19	49.80**
22	words length 5	.01399	.08	.17	.965	1.20	46.96**
23	words length 6	.02668	.14	- .19	.965	1.21	44.36**
24	words length 9	- .03564	.26	- .14	.965	1.21	41.96**

INTERCEPT = 66.815

* significant at .05 level
** significant at .01 level

TABLE 10

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 2 FLUENCY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	# paragraphs	.39813	.19	2.09*	.840	1.50	235.07**
2	# question marks	- .41717	.16	-2.60*	.856	1.43	133.47**
3	words length 9	- .95046	.83	-1.12	.864	1.40	94.60**
4	# words	1.88253	.86	2.19*	.870	1.38	74.23**
5	average sentence length	- .26017	.13	-2.02*	.881	1.33	65.51**
6	SD of sentence length	- .25346	.15	-1.73	.889	1.30	58.25**
7	4th mom. of word length	- .58237	.32	-1.83	.894	1.27	52.44**
8	3rd mom. of word length	.29738	.97	.31	.899	1.25	48.21**
9	SD of word length	-2.02571	1.28	-1.58	.904	1.23	44.52**
10	words length 10	- .63561	.79	- .80	.910	1.20	42.83**
11	word length 4	-1.90348	.88	-2.17*	.913	1.19	40.02**
12	# periods	- .08661	.06	-1.37	.914	1.18	37.01**
13	words length 1	-1.93719	.86	-2.25*	.915	1.18	34.20**
14	words length 2	-1.84587	.86	-2.14*	.917	1.18	31.96**
15	#commas	- .49436	.37	-1.35	.918	1.18	29.94**
16	average word length	-1.08006	1.00	-1.08	.919	1.18	28.09**
17	words length 7	-1.60096	.85	-1.88	.919	1.19	26.38**
18	words length 3	-1.78623	.88	-2.06*	.920	1.19	23.43**
19	# left parentheses	- .69080	1.32	- .53	.921	1.20	22.10**
20	SD of paragraph length	- .34389	1.79	- .19	.921	1.20	22.10**
21	words length 5	-1.82108	.86	-2.00*	.921	1.20	20.86**
22	words length 6	-1.70898	.88	-1.95	.921	1.21	19.69**
23	words length 8	-1.62998	.85	-1.92	.925	1.19	19.69**
24	average paragraph length	- .12791	.67	- .19	.925	1.19	18.63**

INTERCEPT = 11.246

* significant at .05 level
** significant at .01 level

TABLE 11

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 3 FLUENCY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	# paragraphs	.11581	1.86	.08	.923	1.23	562.38**
2	words length 8	- .12393	.43	.29	.932	1.21	321.95**
3	words length 2	.23506	.46	.52	.937	1.17	231.92**
4	words length 7	.06615	.47	.14	.940	1.15	180.93**
5	words length 9	.49586	.43	1.14	.942	1.14	148.18**
6	words length 3	.29026	.45	.59	.944	1.13	126.07**
7	# commas	- .69266	.41	-1.70	.945	1.12	110.55**
8	# periods	.02673	.05	.60	.946	1.12	96.95**
9	words length 5	.31883	.45	.71	.947	1.12	86.12**
10	average sentence length	- .08870	.09	- .95	.947	1.12	77.83**
11	SD of word length	1.49669	1.35	1.11	.948	1.12	70.72**
12	words length 10	- .34502	.63	- .59	.948	1.12	64.63**
13	words length 6	.35531	.46	.77	.949	1.12	59.36**
14	average word length	-1.41060	1.17	-1.20	.949	1.12	55.30**
15	4th mom. of word length	.32683	.29	1.12	.950	1.13	51.71**
16	SD sentence length	.06038	.11	.56	.950	1.13	48.10**
17	words length 1	.11349	.46	.24	.950	1.13	44.86**
18	3rd mom. of word length	- .59008	.81	- .73	.950	1.14	42.09**
19	SD paragraph length	-1.21909	2.70	- .45	.951	1.14	39.46**
20	# sentences	.57668	1.88	.30	.951	1.15	37.07**
21	# question marks	- .09042	.32	- .28	.951	1.16	34.89**
22	average paragraph length	- .38122	1.96	- .19	.951	1.16	32.90**
23	# words	- .22140	.45	- .49	.951	1.17	31.07**
24	words length 4	.22087	.46	.48	.951	1.18	29.48**

INTERCEPT = 3.235

** significant at .01 level

The results of the regression analyses for the Flexibility criteria of Activities 1, 2, and 3 are represented in Tables 12, 13, and 14, respectively. Unexpectedly, the number of responses made by the subjects are found to be the best predictor of Flexibility for Activities 1 and 2. As was hypothesized, "category counts," which were determined from the dictionary constructed according to the guidelines set forth by Torrance, were also found to be of great value as predictors. For Activity 3, Flexibility, "number of words" was the first predictor extracted in the step-wise analysis. At first glance this result might appear to be discordant with the Flexibility result for Activities 1 and 2; however, since the intercorrelations among the variables are high (for these data, correlations of .83 were found to exist between "number of words" and both "number of sentences" and "number of paragraphs"), the "number of words" used can also be considered a measure of the number of responses given. For Activity 3, it was expected that the variable "category counts" would be a better predictor of Flexibility than it was actually found to be. However, as with Fluency, the multiple-R coefficients were higher than had been anticipated. This will be discussed in more detail in Chapter VI.

Tables 15, 16, and 17 summarize the results of the Originality criteria of Activities 1, 2, and 3, respectivel. "Category counts" was the best predictor of Originality for Activity 1, but for Activities 2 and 3 it was not as effective as had been hypothesized. The multiple-R coefficients (.93 for Activity 1, .91 for Activity 2, and .83 for Activity 3) were again higher than anticipated.

Cross-Validation of Prediction Equations

To cross-validate these results the prediction equations derived from the analyses of the developmental data were applied to a new set of data for which N was 53. That is, for each of the nine criteria the established b-weights were used to predict the scores assigned by the human judges. These two sets of scores, the scores assigned by the human judges and the scores assigned by the computer, were then compared using the Pearson Product-Moment Correlation. The correlations thereby obtained are presented in Table 18. The correlation coefficients corrected for attenuation caused by criterion unreliability are also reported in the table. Highly statistically significant correlations were found for all of the nine predicted scores. Moreover, significant increases in the correlations were found when the correction for attenuation was made. The reasons for the implications of these results will be discussed in Chapter VI.

TABLE 12

STEP-WIDE MULTIPLE REGRESSION

ACTIVITY 1 FLEXIBILITY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	# sentences	-34.25142	17.56	-1.95	.781	1.53	153.57**
2	# periods	- .08632	.09	- .94	.834	1.36	110.87**
3	category counts	.45081	.11	-1.10**	.853	1.30	85.69**
4	words length 8	2.26109	2.12	1.06	.869	1.23	73.50**
5	# commas	.87165	.27	3.26**	.875	1.21	61.25**
6	SD of sentence length	- .54947	.22	-2.48*	.880	1.20	53.22**
7	# words	- 3.54497	2.15	-1.65	.885	1.18	47.37**
8	words length 7	3.41739	2.18	1.57	.890	1.16	43.21**
9	# paragraphs	34.30252	17.58	1.95	.893	1.15	39.59**
10	# question marks	.12757	.07	1.90	.897	1.14	36.74**
11	words length 1	3.18281	2.11	1.51	.902	1.12	32.24**
12	SD paragraph length	95.39752	50.90	1.87	.904	1.12	32.24**
13	SD word length	1.98246	1.34	1.48	.905	1.12	29.82**
14	4th mom. of word length	.29580	.37	.81	.906	1.12	27.76**
15	average word length	- 1.41501	1.16	-1.22	.907	1.12	25.84**
16	words length 2	3.58644	2.15	1.67	.907	1.12	24.13**
17	average sentence length	- .04437	.23	- .19	.908	1.13	22.54**
18	words length 5	3.61413	2.16	1.68	.908	1.13	21.14**
19	words length 10	3.51945	2.02	1.74	.908	1.14	19.81**
20	words length 3	3.63082	2.15	1.69	.908	1.14	18.62**
21	words length 6	3.71283	2.17	1.71	.908	1.15	17.53**
22	words length 4	3.65477	2.15	1.70	.908	1.16	16.52**
23	words length 9	3.51562	2.07	1.69	.911	1.18	16.17**
24	3rd mom. word length	- 1.16901	1.52	- .76	.912	1.15	15.44**

INTERCEPT = 2.683

* significant at .05 level
** significant at .01 level

TABLE 13

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 2 FLEXIBILITY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	# paragraphs	-2.62630	4.12	- .64	.742	1.12	102.16**
2	category counts	.31449	.08	3.91**	.796	1.01	83.70**
3	words length 4	.71523	.69	-1.03	.815	.98	63.27**
4	words length 5	- .49762	.68	- .79	.826	.95	50.82**
5	average word length	- .98768	.81	-1.22	.834	.94	43.12**
6	4th mom. word length	- .22482	.25	- .89	.843	.92	38.21**
7	words length 1	- .76081	.68	-1.11	.854	.90	35.29**
8	# question marks	- .16027	.13	-1.26	.858	.89	31.76**
9	words length 3	- .58226	.68	- .85	.860	.89	28.46**
10	SD sentence length	- .11579	.12	- .99	.862	.89	25.77**
11	# periods	.03069	.05	.61	.863	.89	23.40**
12	words length 6	- .52524	.69	- .76	.865	.89	21.47**
13	# commas	- .22491	.29	- .78	.865	.89	19.73**
14	words length 10	- .33807	.64	- .53	.866	.90	18.22**
15	words length 8	- .51794	.67	- .77	.86[illegible]	.90	16.89**
16	words length 2	- .62112	.68	- .92	.867	.90	15.72**
17	# words	.60499	.67	.90	.867	.91	14.67**
18	average sentence length	- .01979	.10	- .19	.868	.92	12.86**
19	average paragraph length	-3.05410	4.07	.75	.868	.92	12.86**
20	# sentences	2.96434	4.14	.72	.868	.92	12.11**
21	SD paragraph length	-3.52642	5.81	- .61	.869	.92	11.46**
22	words length 9	- .56418	.65	- .87	.869	.93	10.81**
23	words length 7	- .56196	.67	- .84	.870	.93	10.32**
24	SD word length	- .13041	1.03	- .17	.870	.94	9.77**
25	3rd mom. word length	- .04188	.78	- .05	.870	.94	9.26**

INTERCEPT = 8.352

** significant at .01 level

TABLE 14

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 3 FLEXIBILITY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	# words	- .34769	.40	- .87	.735	1.16	115.22**
2	average sentence length	- .09212	.08	-1.12	.769	1.10	70.24**
3	words length 10	- .44281	.55	- .80	.783	1.07	50.82**
4	words length 8	.06574	.38	.17	.797	1.05	41.45**
5	words length 9	.68210	.38	1.29	.806	1.03	34.84**
6	words length 7	.24286	.42	.58	.815	1.02	30.57**
7	words length 1	.17655	.40	.44	.825	1.00	27.91**
8	average paragraph length	-1.24916	1.71	- .72	.830	.99	25.13**
9	words length 4	.33906	.40	.84	.834	.99	22.78**
10	# sentences	.52144	1.64	.31	.839	.98	21.09**
11	# commas	- .44338	.36	-1.24	.843	.97	19.66**
12	category counts	.09717	.08	1.17	.846	.97	18.22**
13	# periods	.02642	.04	.64	.847	.97	16.85**
14	4th mom. word length	.24138	.25	.95	.849	.97	15.63**
15	# question marks	.12893	.28	.45	.849	.98	14.51**
16	SD sentence length	.03655	.09	.39	.850	.98	13.49**
17	3rd mom. word length	- .50609	.71	- .71	.850	.99	12.57**
18	SD word length	.79048	1.18	.66	.850	.99	11.75**
19	words length 3	.42220	.40	1.06	.851	1.00	11.01**
20	words length 2	.41219	.40	1.03	.851	1.00	10.35**
21	words length 5	.39992	.40	1.00	.851	1.01	9.76**
22	words length 6	.38698	.40	.95	.853	1.01	9.35**
23	average word length	- .18779	1.05	- .17	.853	1.01	8.83**
24	SD paragraph length	- .65827	2.36	- .27	.853	1.02	8.36**
25	# paragraphs	- .34601	1.62	- .21	.853	1.03	7.92**

INTERCEPT = 1.701

** significant at .01 level

TABLE 15

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 1 ORIGINALITY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	category counts	- .38962	.07	5.25**	.812	1.99	189.61**
2	# question marks	.12775	.08	1.55	.874	1.67	156.54**
3	# paragraphs	3.12315	20.52	.15	.888	1.58	119.46**
4	words length 5	- .60086	2.58	- .23	.893	1.56	94.01**
5	words length 6	- .59173	2.60	- .22	.898	1.53	78.75**
6	SD sentence length	- .65800	.27	-2.42*	.905	1.49	69.85**
7	words length 8	.36036	2.48	- .14	.912	1.44	64.89**
8	words length 2	- .96862	2.57	- .37	.917	1.41	59.76**
9	# periods	- .16656	.11	-1.55	.920	1.40	54.76**
10	SD paragraph length	5.56945	59.44	.09	.921	1.39	50.06**
11	average sentence length	.43023	.28	1.54	.923	1.39	45.81**
12	SD word length	3.06703	1.65	1.86	.925	1.37	43.12**
13	words length 10	-1.36320	2.42	- .56	.926	1.37	39.70**
14	words length 7	-1.00307	2.61	- .38	.927	1.38	36.86**
15	words length 1	-1.06744	2.52	- .42	.927	1.38	34.40**
16	average word length	-1.14971	1.43	.80	.928	1.38	32.33**
17	3rd mom. word length	.11836	1.88	.06	.930	1.37	30.64**
18	words length 8	-1.18217	2.54	- .46	.930	1.38	28.97**
19	words length 4	- .75715	2.58	- .29	.931	1.38	27.30**
20	# sentences	-2.50125	20.50	- .12	.931	1.39	25.67**
21	4th mom. word length	- .22907	.45	- .50	.931	1.40	24.21**
22	# commas	- .09862	.33	- .29	.931	1.40	22.83**
23	# words	.81891	2.58	.31	.931	1.41	21.56**
24	words length 3	- .81553	2.58	- .31	.931	1.42	20.42**

INTERCEPT = -3.211

* significant at .05 level
** significant at .01 level

TABLE 16

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 2 ORIGINALITY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	# sentences	- .89917	9.26	- .09	.716	2.31	103.06**
2	# words	3.11570	1.56	1.99*	.731	2.27	55.58**
3	SD word length	-4.55652	2.29	-1.98*	.747	2.22	40.40**
4	average sentence length	- .21827	.23	- .96	.757	2.20	31.81**
5	# periods	- .30606	.12	-2.63**	.764	2.22	26.38**
6	category counts	.16351	.09	1.73	.774	2.15	23.19**
7	words length 10	-1.16908	1.45	- .80	.781	2.13	20.57**
8	4th mom. word length	- .50461	.54	- .93	.791	2.10	18.96**
9	# question marks	- .30912	.29	-1.08	.795	2.10	17.14**
10	words length 3	-3.05316	1.57	-1.93	.798	2.09	15.62**
11	SD sentence length	- .37740	.27	-1.39	.802	2.09	14.39**
12	words length 8	-2.50528	1.54	-1.63	.805	2.08	13.34**
13	# commas	-1.00861	.65	-1.54	.808	2.08	12.42**
14	words length 4	-3.15526	1.60	-1.96	.809	2.09	11.51**
15	words length 2	-3.13843	1.57	-2.00*	.811	2.09	10.78**
16	3rd mom. word length	.53010	1.78	.29	.815	2.08	10.27**
17	words length 1	-3.11306	1.59	-1.95	.817	2.09	9.64**
18	words length 5	-2.94180	1.57	-1.87	.819	2.09	9.16**
19	words length 9	-2.76279	1.49	-1.85	.820	2.10	8.62**
20	# paragraphs	1.43642	9.22	.15	.820	2.11	8.11**
21	average word length	- .78812	1.78	- .44	.820	2.12	7.65**
22	average paragraph length	1.28449	9.12	.14	.821	2.13	7.21**
23	words length 7	-2.76995	1.56	-1.77	.821	2.15	6.82**
24	word length 6	-2.80227	1.59	-1.76	.829	2.12	6.85**
25	SD paragraph length	1.51536	12.99	.11	.829	2.13	6.49**

INTERCEPT - 11.43612

* significant at .05 level
** significant at .01 level

TABLE 17

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 3 ORIGINALITY

STEP	VARIABLE ENTERED	b-weight	SE	t-value	Mult-R	SE	F-value
1	# sentences	5.49450	3.06	1.79	.853	2.03	261.28**
2	# words	.60784	.74	.82	.881	1.85	168.17**
3	4th mom. word length	.59900	.46	1.31	.886	1.83	116.38**
4	# periods	.08841	.07	1.19	.890	1.81	90.04**
5	words length 10	-1.92395	1.01	-1.91	.893	1.79	74.03**
6	# question marks	.46935	.51	.92	.896	1.78	62.89**
7	words length 6	- .47340	.75	- .63	.898	1.77	54.49**
8	words length 3	- .48036	.74	- .65	.899	1.77	47.80**
9	words length 1	- .66481	.74	- .89	.900	1.77	42.44**
10	words length 8	- .77641	.69	-1.13	.900	1.78	38.01**
11	SD word length	2.88256	2.12	1.35	.901	1.78	34.56**
12	words length 7	- .71573	.78	- .91	.902	1.78	31.73**
13	# commas	.55678	.64	.86	.903	1.78	29.22**
14	SD sentence length	- .04854	.17	- .28	.904	1.79	27.00**
15	words length 4	- .51498	.75	- .68	.904	1.80	25.05**
16	average word length	-1.82616	1.85	- .98	.904	1.80	23.30**
17	3rd mom. word length	- .78717	1.28	- .61	.906	1.80	22.01**
18	words length 2	- .63213	.74	- .85	.906	1.81	20.70**
19	# paragraphs	-4.91796	3.03	-1.62	.907	1.82	19.44**
20	average paragraph length	-4.71293	3.17	-1.48	.908	1.82	18.45**
21	SD paragraph length	-5.66665	4.32	-1.31	.910	1.81	17.78**
22	words length 9	- .54010	.70	- .76	.910	1.82	16.77**
23	words length 5	- .54963	.74	- .74	.910	1.83	15.94**
24	category counts	- .01822	.06	- .32	.910	1.84	15.11**
25	average sentence length	- .02434	.15	- .16	.910	1.85	14.31**

INTERCEPT = 4.833

** significant at .01 level

TABLE 18

VALIDITY OF MULTIPLE REGRESSION EQUATIONS

IN CROSS-VALIDATION SAMPLE

Criterion	Uncorrected Correlation Coefficient	Correlation Coefficient Corrected for Attenuation
Activity 1, Fluency	.89**	.90**
Acticity 1, Flexibility	.71**	.72**
Activity 1, Originality	.74**	.83**
Activity 2, Fluency	.88**	.90**
Activity 2, Flexibility	.68**	.71**
Activity 2, Originality	.75**	.89**
Activity 3, Fluency	.88**	.92**
Activity 3, Flexibility	.56**	.59**
Activity 3, Originality	.72**	.99**

** significant at .01 level

CHAPTER V

RESULTS — ACTIVITIES 4, 5, 6, AND 7
(by John F. Greene)

INTRODUCTION

In this chapter the results of the computer simulation of the human rating practices for Activities 4, 5, 6, and 7 are reported. The chapter is logically partitioned into two sections; the multiple regression results for the developmental sample and the results obtained in the empirical cross-validation procedures.

Establishing Prediction Equations

Three sets of multiple regression analyses, each one corresponding to a particular model, are considered in this section. The first eleven tables (19 through 29) consist of the full model results for each dimension of Activities 4, 5, 6, and 7. In this model, all appropriate variables are allowed to enter the regression process. The restricted model is based on a subset of predictors statistically selected (i.e. by means of step-wise multiple regression) from the full set, and certain parts of the results are equivalent to the corresponding sections in the full model. The number of predictors, or the size of the partial set, is determined by the investigator and will be discussed at a later point in this chapter. Two forced models are considered in Tables 30 and 31. In this type of analysis, the researcher selects a partial set of predictors and forces them into the analysis before the remaining variables of the full set are allowed to enter. If this model is to differ from the full model, it also must be restricted.

The same format is employed for the presentation of results for each of the models. The numbers corresponding to the variable entered are defined in Figure 7. Variables 1, 2, and 3 represent the criteria or dependent variables for Fluency, Flexibility and Originality, and, of course, will not enter as predictor variables in the analyses. Variables 4 through 24 correspond to the actuarial variables. The category counts for Flexibility and Originality are variables 25 and 26 respectively.

The remaining columns of the tables are partitioned into two fields. The first field contains information relative to the final step of the particular model. Three sets of statistics are given. First, the resultant linear weight or b-weight is presented for each predictor variable. The intercept is also included. Then the standard error of the b-weight is given, followed by a t-value. This t-value is the quotient of the b-weight and the standard error, and provides information pertaining to the value of the coefficient in

TABLE 19

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 4, FLUENCY

FULL MODEL

Step	Variable Entered	Final Step b-weight	Final Step SE	Final Step t-value	Cumulative Mult-R	Cumulative SE	Cumulative F-value
1	18	.75930	.062	12.22**	.975	1.506	1922.68**
2	12	- .18204	.727	- .25	.978	1.429	1073.13**
3	10	- .31105	.740	- .42	.980	1.372	779.26**
4	14	- .67818	.766	- .88	.982	1.312	641.43**
5	4	.48049	.199	2.40*	.983	1.265	553.98**
6	19	- .05467	1.035	- .05	.985	1.226	492.19**
7	22	- .13258	.150	- .87	.985	1.219	427.59**
8	15	.02187	.652	.03	.985	1.213	377.69**
9	16	.52722	.781	.67	.985	1.211	337.05**
10	24	- .13913	.201	- .69	.986	1.205	306.42**
11	7	- .49518	.743	- .66	.986	1.206	278.12**
12	11	- .40057	.738	- .54	.986	1.210	253.57**
13	9	- .35986	.753	- .47	.986	1.214	232.35**
14	21	1.01944	1.548	.65	.986	1.220	213.89**
15	23	- .50435	.847	- .59	.986	1.222	198.87**
16	13	- .40364	.736	- .54	.986	1.229	184.51**
17	20	- .02138	.076	- .27	.986	1.235	171.77**
18	8	- .39426	.754	- .52	.986	1.243	160.33**
19	17	.38283	.742	.51	.986	1.248	150.52**
20	5	- .00346	.024	- .14	.986	1.256	141.25**

INTERCEPT: -.45530

* Significant at .05 level
** Significant at .01 level

TABLE 20

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 4, FLEXIBILITY

FULL MODEL

Step	Variable Entered	Final Step			Cumulative		
		b-weight	SE	t-value	Mult-R	SE	F-value
1	25	.40526	.068	5.88**	.791	1.497	163.87**
2	18	.09021	.056	1.59	.849	1.298	125.63**
3	20	- .25613	.069	-3.69**	.865	1.242	94.76**
4	13	.58905	.687	.85	.873	1.211	76.19**
5	12	.57567	.681	.84	.879	1.191	63.83**
6	7	.16915	.696	.24	.886	1.165	56.50**
7	4	.33770	.181	1.86	.889	1.154	49.72**
8	15	.76220	.602	1.26	.892	1.147	44.39**
9	19	-2.65844	.935	-2.84**	.895	1.137	40.37**
10	8	.38354	.706	.54	.898	1.132	36.90**
11	11	.45183	.691	.65	.900	1.125	34.09**
12	21	3.21705	1.403	2.29*	.902	1.120	31.72**
13	23	-1.28817	.764	-1.68	.907	1.101	30.59**
14	5	.02244	.022	1.01	.908	1.100	28.53**
15	17	- .42414	.695	- .61	.909	1.101	26.64**
16	22	.07536	.136	.55	.910	1.105	24.85**
17	16	.10033	.714	.14	.910	1.109	23.20**
18	14	.50084	.718	.69	.910	1.116	21.69**
19	24	.12963	.184	.70	.910	1.122	20.32**
20	9	.45005	.704	.63	.910	1.128	19.11**
21	10	4.2912	.692	.61	.911	1.132	18.08**

INTERCEPT: 8.36219

* Significant at .05 level
** Significant at .01 level

TABLE 21

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 4, ORIGINALITY

FULL MODEL

Step	Variable Entered	Final Step			Cumulative		
		b-weight	SE	t-value	Mult-R	SE	F-value
1	18	.23268	.132	1.75	.770	2.836	142.40**
2	13	2.43504	1.584	1.53	.794	2.712	82.93**
3	9	2.25689	1.618	1.39	.809	2.637	60.67**
4	12	2.23072	1.564	1.42	.819	2.589	48.34**
5	4	.51945	.429	1.21	.825	2.564	39.98**
6	24	.24407	.430	.56	.826	2.559	33.70**
7	26	.18809	.114	1.64	.831	2.548	29.41**
8	17	-2.10847	1.597	-1.32	.834	2.540	26.07**
9	5	- .02764	.052	- .53	.835	2.549	23.[illegible]5**
10	20	- .11755	.164	- .71	.836	2.559	20.6[illegible]**
11	23	-1.87096	1.812	-1.03	.836	2.570	18.62**
12	21	3.73980	3.323	1.12	.837	2.579	16.97**
13	14	2.01002	1.649	1.21	.837	2.591	15.53**
14	22	.12753	.322	.39	.838	2.603	14.31**
15	7	2.03048	1.599	1.27	.838	2.616	13.22**
16	15	1.86188	1.403	1.32	.838	2.631	12.27**
17	10	2.09022	1.592	1.31	.838	2.646	11.41**
18	8	2.12358	1.623	1.30	.838	2.663	10.65**
19	11	2.07420	1.587	1.30	.840	2.665	10.11**
20	16	1.54420	1.683	.91	.842	2.668	9.63**
21	19	.56489	2.221	.25	.842	2.684	9.06**

INTERCEPT: -6.95818

* Significant at .05 level
** Significant at .01 level

TABLE 22

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 5, FLUENCY

FULL MODEL

Step	Variable Entered	Final Step			Cumulative		
		b-weight	SE	t-value	Mult-R	SE	F-value
1	18	.49678	.079	6.28**	.927	3.083	602.82**
2	12	- .62244	.784	- .79	.938	2.864	357.49**
3	10	- .94048	.755	-1.24	.944	2.754	260.83**
4	21	4.47109	1.343	2.50*	.946	2.712	202.77**
5	19	-2.11771	1.049	-2.01*	.949	2.655	170.20**
6	7	-1.19917	.748	-1.60	.951	2.617	146.63**
7	15	-1.41491	.792	-1.78	.952	2.596	128.05**
8	11	- .91738	.784	-1.17	.954	2.566	115.07**
9	20	- .81247	.357	-2.27*	.956	2.524	106.15**
10	9	-1.03074	.753	-1.36	.957	2.498	97.85**
11	23	.54817	1.216	.45	.958	2.502	88.77**
12	17	1.09254	.753	1.44	.958	2.509	80.97**
13	4	- .10873	.379	- .28	.958	2.518	74.21**
14	13	-1.18954	.772	-1.54	.958	2.530	68.24**
15	8	-1.12172	.760	-1.47	.958	2.542	63.16**
16	14	-1.18653	.815	-1.45	.959	2.542	59.25**
17	16	- .64605	.803	- .80	.959	2.547	55.57**
18	24	- .20930	.391	- .53	.959	2.559	52.00**
19	5	.03329	.050	.65	.960	2.571	48.85**
20	22	.19216	.365	.52	.960	2.582	46.00**

INTERCEPT: 6.19578

* Significant at .05 level
** Significant at .01 level

TABLE 23

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 5, FLEXIBILITY

FULL MODEL

Step	Variable Entered	Final Step			Cumulative		
		b-weight	SE	t-value	Mult-R	SE	F-value
1	25	.56353	.096	5.85**	.767	2.114	140.22**
2	12	.08259	.602	.13	.790	2.032	80.43**
3	10	.02008	.582	.03	.811	1.950	61.28**
4	20	- .39275	.273	-1.43	.828	1.876	51.86**
5	9	- .04580	.581	- .07	.832	1.868	42.22**
6	15	- .29450	.614	- .47	.834	1.867	35.41**
7	11	.02897	.603	.04	.838	1.855	31.02**
8	14	- .22230	.627	- .35	.839	1.861	27.05**
9	19	.82473	.801	1.02	.840	1.866	23.97**
10	7	- .00978	.577	- .01	.841	1.869	21.55**
11	4	.22635	.293	.77	.843	1.872	19.61**
12	16	- .30126	.613	- .49	.844	1.878	17.89**
13	8	- .03629	.586	- .06	.844	1.887	16.38**
14	18	- .02277	.062	- .36	.845	1.895	15.10**
15	24	.14733	.300	.49	.845	1.905	13.96**
16	23	- .47550	.929	- .51	.845	1.914	12.98**
17	21	.25285	1.033	.24	.845	1.925	12.08**
18	22	.07523	.283	.26	.845	1.936	11.18**
19	13	- .07702	.593	- .13	.846	1.947	10.56**
20	17	.04565	.581	.07	.846	1.959	9.91**
21	5	.00256	.038	.06	.846	1.972	9.32**

INTERCEPT: -.31133

** Significant at .01 level

TABLE 24

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 5, ORIGINALITY

FULL MODEL

Step	Variable Entered	Final Step			Cumulative		
		b-weight	SE	t-value	Mult-R	SE	F-value
1	18	.20273	.085	2.37*	.718	3.464	104.03**
2	26	.20433	.064	3.18**	.793	3.048	81.99**
3	8	.18933	.823	.23	.814	2.916	62.99**
4	22	.97045	.396	2.45*	.834	2.788	54.20**
5	16	2.04265	.871	2.34*	.849	2.681	48.67**
6	24	- .39889	.425	- .93	.853	2.662	41.50**
7	5	.05104	.055	.91	.857	2.643	36.42**
8	7	.24205	.810	.29	.859	2.644	31.96**
9	15	.42055	.857	.49	.860	2.648	28.40**
10	12	.44884	.848	.52	.861	2.652	25.56**
11	10	.34877	.817	.42	.862	2.659	23.16**
12	23	.43820	1.320	.33	.863	2.671	21.06**
13	11	.37825	.848	.44	.863	2.684	19.27**
14	20	- .35887	.389	- .92	.863	2.692	17.80**
15	19	- .65828	1.173	- .56	.864	2.705	16.49**
16	21	1.28958	1.486	.86	.865	2.714	15.37**
17	9	.31778	.815	.38	.865	2.727	14.35**
18	17	- .29322	.816	- .35	.865	2.741	13.42**
19	14	.25517	.882	.28	.865	2.758	12.56**
20	13	.23523	.835	.28	.866	2.774	11.80**
21	4	.00971	.413	.02	.866	2.792	11.09**

INTERCEPT: .93614

* Significant at .05 level
** Significant at .01 level

TABLE 25

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 6, FLUENCY

FULL MODEL

Step	Variable Entered	Final Step b-weight	Final Step SE	Final Step t-value	Cumulative Mult-R	Cumulative SE	Cumulative F-value
1	10	- .46266	.643	- .71	.846	2.323	245.96**
2	9	- .41179	.644	- .63	.923	1.683	279.07**
3	7	-1.11339	.648	-1.71	.945	1.437	267.57**
4	18	.45919	.099	4.63**	.957	1.276	261.22**
5	5	- .06037	.067	- .89	.963	1.194	241.38**
6	17	.57828	.645	.89	.966	1.148	219.13**
7	6	.04971	.038	1.29	.967	1.134	188.97**
8	23	-1.34250	.581	-2.30*	.967	1.144	166.01**
9	24	.29896	.169	1.76	.969	1.126	152.77**
10	12	- .49586	.694	- .71	.969	1.123	138.24**
11	21	1.43612	.596	2.40*	.970	1.124	125.63**
12	19	- .66263	.307	-2.15*	.971	1.096	121.55**
13	16	-1.64787	1.188	-1.38	.972	1.095	112.42**
14	11	- .51477	.653	- .78	.972	1.098	103.97**
15	22	.13333	.168	.79	.972	1.101	96.41**
16	20	- .04715	.119	- .39	.972	1.106	89.60**
17	14	- .63113	.659	- .95	.972	1.112	83.47**
18	15	- .60902	.670	- .90	.972	1.118	77.99**
19	8	- .56363	.645	- .87	.972	1.124	73.06**
20	13	- .57648	.703	- .82	.973	1.126	69.19**
21	4	.01238	.325	.03	.973	1.134	65.06**

INTERCEPT: .23444

* Significant at .05 level
** Significant at .01 level

TABLE 26

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 6, ORIGINALITY

FULL MODEL

Step	Variable Entered	Final Step b-weight	Final Step SE	Final Step t-value	Cumulative Mult-R	Cumulative SE	Cumulative F-value
1	9	.28572	1.368	.20	.625	2.768	62.67**
2	26	.22038	.079	2.77**	.661	2.673	37.67**
3	13	1.14729	1.485	.77	.692	2.584	29.46**
4	24	.31147	.359	.86	.723	2.487	26.00**
5	16	7.55586	2.502	3.01**	.744	2.418	23.31**
6	14	.88244	1.391	.63	.755	2.384	20.61**
7	23	- .86116	1.228	- .70	.762	2.370	18.17**
8	18	- .66736	.208	-3.19**	.767	2.361	16.24**
9	10	.27155	1.358	.27	.776	2.334	15.12**
10	5	.25710	.141	1.81	.790	2.281	14.77**
11	6	.09347	.080	1.15	.794	2.273	13.66**
12	7	.34057	1.371	.24	.797	2.273	12.61**
13	20	- .32550	.251	-1.29	.800	2.272	11.73**
14	19	.34033	.648	.52	.801	2.281	10.84**
15	8	.20396	1.369	.14	.802	2.288	10.09**
16	12	.26793	1.470	.18	.803	2.297	9.39**
17	4	- .14504	.683	- .21	.803	2.310	8.75**
18	15	.25022	1.414	- .17	.803	2.324	8.17**
19	17	- .14529	1.366	- .10	.803	2.337	7.65**
20	11	.11138	1.381	.08	.803	2.352	7.18**
21	22	.02719	.363	.07	.803	2.367	6.75**
22	21	- .04603	1.260	- .03	.803	2.382	6.36**

INTERCEPT: -.10061

* Significant at .05 level
** Significant at .01 level

TABLE 27

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 7, FLUENCY

FULL MODEL

Step	Variable Entered	Final Step			Cumulative		
		b-weight	SE	t-value	Mult-R	SE	F-value
1	17	-1.03187	.739	-1.39	.820	1.758	201.84**
2	8	.91242	.749	1.21	.845	1.651	121.44**
3	20	- .22522	.103	-2.17*	.872	1.522	101.37**
4	11	1.29391	.753	1.71	.884	1.458	85.23**
5	13	1.39626	.745	1.87	.889	1.435	71.23**
6	18	.13342	.112	1.18	.895	1.409	62.29**
7	7	.83857	.733	1.14	.899	1.390	55.34**
8	9	1.11614	.741	1.50	.902	1.377	49.72**
9	5	- .11061	.060	-1.82	.905	1.362	45.50**
10	12	1.10765	.753	1.47	.908	1.354	41.64**
11	22	- .25051	.157	-1.59	.911	1.342	38.80**
12	21	2.16537	1.163	1.86	.912	1.339	35.86**
13	15	.74136	.758	.97	.914	1.332	33.58**
14	19	- .57680	.490	-1.17	.914	1.336	31.01**
15	23	- .67102	.568	-1.18	.915	1.340	28.81**
16	24	.36174	.221	1.63	.916	1.339	27.10**
17	4	.09996	.251	.39	.917	1.344	25.34**
18	14	1.25221	.737	1.69	.917	1.352	23.68**
19	10	1.23943	.745	1.66	.917	1.355	22.36**
20	16	1.26361	.867	1.45	.920	1.345	21.65**

INTERCEPT: .35932

* Significant at .05 level
** Significant at .01 level

TABLE 28

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 7, FLEXIBILITY

FULL MODEL

Step	Variable Entered	Final Step			Cumulative		
		b-weight	SE	t-value	Mult-R	SE	F-value
1	25	.27692	.076	3.60**	.667	1.197	78.44**
2	9	.13216	.548	.24	.748	1.071	61.58**
3	10	.07659	.553	.13	.770	1.036	46.49**
4	15	.36247	.557	.65	.779	1.022	36.75**
5	20	- .23220	.076	-3.05**	.786	1.014	30.37**
6	18	- .15744	.082	-1.91	.801	.986	27.84**
7	21	1.55484	.856	1.81	.812	.965	25.53**
8	16	.56354	.648	.86	.820	.953	23.36**
9	23	- .75189	.415	-1.80	.828	.940	21.75**
10	7	- .17959	.541	- .33	.831	.938	19.83**
11	14	- .16401	.543	- .30	.833	.938	18.11**
12	5	.03979	.044	.89	.835	.939	16.66**
13	13	- .01580	.554	- .02	.836	.942	15.30**
14	17	.05402	.547	.09	.836	.945	14.13**
15	19	- .20322	.358	- .56	.837	.948	13.14**
16	22	.04148	.115	.36	.838	.953	12.20**
17	24	.02677	.162	.16	.838	.959	11.35**
18	8	- .06527	.553	- .11	.838	.964	10.60**
19	11	- .07774	.557	- .13	.838	.970	9.91**
20	12	- .07488	.558	- .13	.838	.976	9.30**
21	4	- .00762	.185	- .04	.838	.982	8.75**

INTERCEPT: -.00051

** Significant at .05 level

TABLE 29

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 7, ORIGINALITY

FULL MODEL

Step	Variable Entered	Final Step b-weight	Final Step SE	Final Step t-value	Cumulative Mult-R	Cumulative SE	Cumulative F-value
1	17	.44306	.450	.98	.526	.900	37.40**
2	8	- .46037	.455	-1.01	.590	.858	25.87**
3	26	.23826	.110	2.14*	.612	.845	19.14**
4	18	- .14725	.071	-2.05*	.631	.834	15.69**
5	16	.43075	.526	.81	.653	.818	13.99**
6	20	- .12448	.062	-1.97	.665	.811	12.29**
7	11	-1.46145	.458	-1.00	.678	.802	11.18**
8	19	.13693	.308	.44	.696	.788	10.67**
9	4	- .09376	.152	- .61	[illegible]	.788	9.62**
10	24	- .13371	.136	- .98	.703	.789	8.70**
11	12	- .40779	.458	- .88	.709	.787	8.08**
12	14	- .48627	.447	-1.08	.712	.789	7.43**
13	15	- .11823	.458	- .25	.716	.788	6.98**
14	10	- .35857	.454	- .78	.719	.789	6.50**
15	23	- .35264	.345	-1.02	.720	.792	6.04**
16	21	.50741	.739	.68	.722	.795	5.65**
17	7	- .38936	.444	- .87	.724	.790	5.31**
18	5	.01883	.036	.51	.725	.802	4.98**
19	22	- .00709	.095	- .07	.725	.806	4.66**
20	13	- .34173	.453	- .75	.725	.811	4.38**
21	9	- .32959	.450	- .73	.727	.814	4.17**

INTERCEPT: .08230

* Significant at .05 level

** Significant at .01 level

TABLE 30

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 7, FLEXIBILITY

FORCED MODEL

Step	Variable Entered	Final Step b-weight	Final Step SE	Final Step t-value	Cumulative Mult-R	Cumulative SE	Cumulative F-value
1	25	.35853	.076	4.66**	.667	1.197	78.44**
2	10	.10137	.025	4.01**	.720	1.121	52.09**
3	21	.42536	.290	1.46	.727	1.114	35.85**

INTERCEPT: -.46068

** Significant at .05 level

TABLE 31

STEP-WISE MULTIPLE REGRESSION

ACTIVITY 7, ORIGINALITY

FORCED MODEL

Step	Variable Entered	Final Step b-weight	Final Step SE	Final Step t-value	Cumulative Mult-R	Cumulative SE	Cumulative F-value
1	10	.05084	.019	2.67**	.467	.935	27.27**
2	21	.56387	.219	2.56*	.541	.894	20.07**
3	9	.05691	.020	2.77**	.586	.866	16.77**
4	26	.15999	.100	1.59	.601	.859	13.42**

INTERCEPT: -.57268

* Significant at .05 level
** Significant at .01 level

Variable Number	Variable Name
1	Fluency criterion
2	Flexibility criterion
3	Originality criterion
4	Number of commas (,)
5	Number of Periods (.)
6	Number of question marks (?)
7	Words of length 1
8	Words of length 2
9	Words of length 3
10	Words of length 4
11	Words of length 5
12	Words of length 6
13	Words of length 7
14	Words of length 8
15	Words of length 9
16	Words of length 10
17	Total number of words
18	Total number of sentences
19	Mean word length
20	Mean sentence length
21	Standard deviation of word length
22	Standard deviation of sentence length
23	Third moment of word length
24	Fourth moment of word length
25	Flexibility Dictionary
26	Originality Dictionary

FIGURE 7

VARIABLE NUMBERS AND NAMES

the population. The null hypothesis is that the beta-weight, or regression coefficient in the population, does not differ from zero.

The second field also contains three sets of statistics for each step; namely, the multiple regression coefficient, the standard error of the estimate, and an F-value. The square of the multiple regression coefficient yields the percentage of variance accounted for in the criterion, and is a good indicator of how well we can predict. The standard error of the estimate is equivalent to the standard deviation of the residuals, or differences between the predicted score and the observed score. The F-value indicates how well a particular set of predictors is able to estimate the criterion. Because the most useful predictors are statistically selected first, the F-value is expected to decrease as predictors are added. Note, however, that at the same time the multiple correlation is increasing.

Excellent results were obtained for Activity 4 and Activity 5. The full model multiple regression coefficients for Fluency, Flexibility, and Originality are .99, .91, and .84 in Activity 4 and .96, .85, and .87 in Activity 5. Equally encouraging results were noted for Activity 6, where the values of .97 and .80 were determined for Fluency and Originality respectively. Once again, the reader is reminded that there is no Flexibility score for Activity 6. Moreover, only minimal differences were found between the full and restricted model results for these activities. In no case did this difference exceed .02.

Somewhat less encouraging results were generated for Activity 7, where the multiple regression coefficients for the full model were .92, .84, and .73. These values dropped to .90, .83, and .70 in the restricted model. The results of the forced model for the Flexibility and Originality dimensions of this activity were .73 and .60. This forced model was only produced after the other two models generated equations which did not predict the criterion scores in the cross-validation sample to the high degree desired, as will soon be shown.

Cross-Validation of Prediction Equations

The cross-validation correlations appear in Table 32. Because of criterion unreliability, an estimate corrected for attenuation is also presented. In an effort to facilitate comparisons, the results of all three types of models are included, as well as the corresponding multiple correlations obtained from the last step in each of the multiple regression tables.

As can be seen by analyzing the results presented in the Table, the attenuated cross-validation correlation coefficients for Activities 4, 5, and 6 are very high for both the full and restricted models. The Fluency range for these Activities is defined by .93 and .96 for the full model, while the values in the restricted model

TABLE 32

CROSS-VALIDATION RESULTS

INCLUDING MULT-R SUMMARY

ALL MODELS*

Activity	Dimension	Full Model			Restricted Model			Forced Model		
		R	r	r attenuated	R	r	r attenuated	R	r	r attenuated
4	Fluency	.99	.95	.96	.99	.95	.96			
4	Flexibility	.91	.80	.91	.90	.84	.86			
4	Originality	.84	.74	.79	.84	.78	.83			
5	Fluency	.96	.92	.93	.95	.95	.96			
5	Flexibility	.85	.87	.89	.84	.89	.91			
5	Originality	.87	.80	.88	.86	.82	.90			
6	Fluency	.97	.95	.95	.97	.96	.96			
6	Originality	.80	.68	.92	.79	.70	.95			
7	Fluency	.92	.84	.87	.90	.82	.85			
7	Flexibility	.84	.53	.56	.83	.56	.59	.73	.73	.77
7	Originality	.73	.42	.48	.70	.55	.62	.60	.62	.70

* All correlations are significant at .01 level

did not vary from .96. The Flexibility coefficients were .81 and .89 in the full model and .86 and .91 in the restricted model for Activities 4 and 5. Originality varies from .79 to .92 in the full model as compared to .83 to .95 in the restricted model. Only for the dimension of Originality were considerable differences noted between the unattenutated and attenuated correlations. This was expected, however, since the pooled reliabilities of the judges for this dimension was lower than Fluency and Flexibility.

In Activity 7 the full and restricted model attenuated cross-validation correlations were .87 and .85 for Fluency, .56 and .59 for Flexibility, and .48 and .62 for Originality. Although these results indicated that the prediction equations are useful, they are somewhat lower than the results obtained for Activities 4 through 6. Thus, a forced model was generated for Flexibility and Originality. Correlations of .77 and .70 were obtained when the forced model prediction equations were compared to the observed scores.

Summary

The results of the extensive statistical analyses of this study were presented in Chapter IV and V. Included were several multiple correlation analyses, which were used to generate prediction equations, and a cross-validation approach to the evaluation of the entire computerized scoring procedure. The discussion of these results along with their implications for future research will be presented in the next chapter.

CHAPTER VI

DISCUSSION OF RESULTS

This chapter will describe some of the findings, and the implications of the findings from the attempt to predict scores on the Torrance Tests of Creative Thinking by means of computer simulation.

Prediction Equations — Activities 1, 2, and 3

In multivariate analysis, it is often pointless to elaborate a hypothesis for each predictor, or to explain how each variable met or failed to meet expectations. Moreover, since many of the predictors included in this research were "variables of opportunity," that is, variables which had been shown in previous research to be of some value in content analysis but not for the content analysis of responses to creativity tests, strong relationships were not expected in all cases. In particular, the ten simple word length statistics (number of words of length one, two, three, etc.) were expected to aid in the prediction of creativity scores, but they were not expected to be among the best predictors. As is evident from the data presented in Chapters IV and V, however, in many cases these simple counts were extracted early in the step-wise multiple regression analyses. Although the variables "number of sentences," "number of paragraphs," and "number of words" were hypothesized and observed to be among the best predictors of Fluency, an unexpected finding was that the same variables were also found to be among the best predictors of Flexibility and Originality.

For the prediction of Flexibility and Originality, it was hypothesized that the variable "category counts" would be the most important predictor, since the counts were derived from the dictionaries in accordance with Torrance's scoring norms. However, this was true only for the prediction of the Activity 1, Originality scores. For the prediction of the Flexibility scores of Activities 1 and 2, "category counts" was the sixth best predictor; for Activity 3, Flexibility, it was the twelfth best predictor; and for Activity 3, Originality the variable was not entered until the 24th step of the regression analysis. That "category counts" was not more important than had been hypothesized could have been the result of insufficient or incorrect entries in the dictionaries or categories. But if the correlations between "category counts" and the criteria were indeed high, the value of "category counts" as a predictor would have been influenced by: a) high correlations between other variables and the criteria, and b) high correlations between these other variables and "category counts." If this was the case, much of the variance in the criterion which could be accounted for by "category counts" would be extracted by the other variables. To investigate this possibility, the simple correlations among the predictors themselves and the correlations between the "category counts" and the various criteria were examined. Table 33 reports the Pearson Product-Moment Correlation Coefficients for the criteria and the various "category counts." Due to the voluminous nature of the intercorrelations among the variables, no such tables have been provided here.

As indicated in Table 33, the correlations of the "category counts" with the criteria are generally very high, ranging from .54 to .81. It should be noted that the size of the correlations with the criteria does not necessarily indicate the importance of "category counts" as a predictor. For example, for Acticity 1, Flexibility, "category counts," which correlated .74 with its criterion, was introduced third in its step-wise regression analysis. However, although the correlation was .67, the same variable was introduced in the second step when Flexibility was being predicted for Activity 3. When the intercorrelations among the predictors are considered, the reason for this is apparent. In all cases where "category counts" was not the best predictor, a high correlation was found between "category counts" and one of the best predictors. For example, when the Flexibility score for Activity 3 was sought, the best predictor, "number of words," correlated .78 with "category counts." It seems, then, that the dictionaries or categories were at least adequate, but that their importance was restricted in the various prediction equations by the importance of other predictors.

As indicated, the number of responses, as measured by the variables "number of sentences," "number of paragraphs," and "number of words," were continually among the best predictors of each of the three dimensions of creativity. Since these three variables are all measures of verbal Fluency, as Fluency is understood in a literary sense rather than as used by Torrance, it is possible that there exists one underlying dimension for the TTCT, rather than the three dimensions, Fluency, Flexibility, and Originality. Other evidence for this interpretation has been reported in the literature. Cicirelli (1964), for example, reported intercorrelations of .79 for Fluency and Flexibility, .80 for Fluency and Originality, and .74 for Originality and Flexibility. Long and Henderson (1964) have found average intercorrelations for samples of children in grades two through seven of .68 for Fluency and Flexibility, .60 for Fluency and Originality, and .80 for Flexibility and Originality. The interpretation given these results was that subjects high or low in Fluency, as measured by the TTCT, would likewise be high or low in Flexibility and Originality. Or again, highly creative people would simultaneously be Fluent, Flexible, and Original while those of low creativity would simultaneously be less Fluent, Flexible, and Original. However, the results reported here suggest that persons found to be creative by the TTCT are highly Fluent and that Fluency accounts for their high Flexibility and Originality scores. Likewise, those persons who are not Fluent, that is, do not give many responses, also will not be found to be Flexible and Original. On common sense grounds, it would appear that one improves his chances of increasing his Flexibility score simply by producing a greater number of

TABLE 33

SIMPLE CORRELATIONS BETWEEN "CATEGORY COUNTS" AND CRITERIA

Criteria	r	Step Variable Entered into Regression Equation
Activity 1, Flexibility	.74**	3
Activity 1, Originality	.81**	1
Activity 2, Flexibility	.67**	2
Activity 2, Originality	.54**	6
Activity 3, Flexibility	.60**	12
Activity 3, Originality	.59**	24

** Significant at .05 level

responses. And, of course, chances of "scoring a hit" in Originality are likely to increase when one generates a greater number of responses. Whether this relationship does, in fact, exist must be determined by further research. At least two studies dealing with the dimensionality issue and growing out of the research were reported by Paulus (1970) and Renzulli (1970) and it is anticipated that further research in the area of creativity must take account of the important problems related to dimensionality.

The Overall Predictive Value of the Scoring Strategy for Activities 1, 2, and 3

From the standpoint of overall simulation, the multiple correlations obtained for the prediction of the pooled human judgments were the primary goal of the analyses. For each of the nine prediction equations derived, the results obtained were rather startling. High multiple correlations were expected for the predictioñ the Fluency dimension of creativity, but the multiple-R's of .97, .93, and .95 obtained for Activities 1, 2, and 3 respectively were higher than anticipated. It was hypothesized that the prediction of the Flexibility dimension of creativity would be a harder task than the prediction of Fluency. The multiple-R's of .91, .87, and .85 for the Flexibility scores of Activities 1, 2, and 3 substantiate this hypothesis. However, the prediction of Flexibility is still very good, since much of the variance in the Flexibility criteria can be accounted for by the set of predictors employed.

Since the assessment of Originality has been found to be a difficult task for humans, even more difficult than the assessment of Flexibility, it was expected that the lowest multiple-R's would be encountered in the computerized prediction of this dimension of creativity. It is not surprising, then, that the lowest multiple correlation coefficient, .83, was found for Activity 2, Originality. However, in the light of both the hypothesis and this finding, the multiple-R's of .93 and .91 for Activity 1, Originality and Activity 3, Originality were unexpected. That the computer can judge Originality (as defined by Torrance) to the degree observed is indeed an important finding.

It is well known, however, that the accuracy found in the derivation of prediction equations should not be expected if new responses to the TTCT were taken and the discovered b-weightings were applied to them to predict their human ratings. For any set of scores, or any set of resultant correlations, contains not only true variance associated with the variables, but also a certain amount of error variance (probably unique to the particular subjects concerned) which will not ordinarily be found with a new set of human subjects, or responses (Page and Paulus, 1968, p. 53). The true variance gives us information which will be subsequently useful. But the error variance is also capitalized upon by the analysis, and a certain portion of the multiple-regression coefficients, and of the contributing b-weights, will spuriously contribute, but will not stand up in a replication.

When the findings are cross-validated, then, the resulting prediction will not always correlate as highly as one might hope. The statistical loss is commonly spoken of as "shrinkage" and has been widely treated in the literature (e.g., McNemar, 1962). As one would suppose, the larger the number of subjects, the more reliable the multiple-R will be; but the larger the number of predictors (given the same number of subjects), the less reliable the multiple-R will be.

The multiple-R's for the Fluency dimension of creativity cross-validated very well, but sizable shrinkage was found for the multiple-R's of the Flexibility and Originality dimensions. However, when adjustments were made for the lack of perfect reliability in the criteria (i.e., the so-called "correction for attenuation"), significant increases in the correlations were found for both of these dimensions. Since these correlations were influenced by both the sample size and the number of predictors, as discussed above, the results are not unexplainable. Also, as noted in the previous chapters, it appears that accuracy in the prediction of any of the dimensions of creativity would not be significantly reduced by the elimination of some of the predictors. Moreover, if fewer predictors were used, the correlations found in cross-validating the results would have been higher. The reader will note that these considerations were brought to bear on the analyses of Activities 4 through 7.

Prediction Evaluations — Activities 4, 5, 6, and 7

The results of the multiple correlation analyses for Activities 4 through 7 presented in Chapter V, must be considered most encouraging. The full model coefficients for Fluency range from .92 to .99. The range for Flexibility is .84 to .91. And the multiple-R's for Originality are .84, .87, .80 and .73 for Activities 4-7 respectively. Although these results must be validated in the cross-validation sample, they represent very high potential prediction power. The percentage of variance accounted for varies from 53 to 98, and each multiple correlation coefficient is significant beyond the .01 level.

The restricted model results parallel those of the full model. In all but one equation, the multiple correlation coefficient dropped by less than one-hundreth of a point. The greatest loss in potential predictability was realized in Activity 7, Originality, where a .03 difference was noted. In these restricted analyses, no more than half of the original set of predictors was utilized, with four instances of using a few as five or six predictors.

Greater losses in the multiple-R coefficient were detected for the two forced models. Activity 7, Flexibility dropped from .84 to

.73, and a .13 decline from .73 was noted for the Activity 7, Originality forced model. These models were generated, however, because of low cross-validations in their respective full and restricted models, as will be shown below. Thus, while lower multiple correlations were obtained, higher cross-validation correlations are expected. The procedure for selecting which predictors were to be forced will be explained in the next section, but one advantage of the particular forced models considered is that they employ only three and four predictors.

All of the multiple correlation coefficients reported are high and significant beyond the .01 level. Before speculating on the relative value of these results, however, the validity of the prediction equations will be estimated in the next section.

Cross-Validaton of Prediction Equations (Activities 4 through 7)

The attenuated cross-validation correlations appear in Table 32. Only the first nine equations will be immediately discussed, followed by the equations relative to the Flexibility and Originality scores for Activity 7. The cross-validation correlations for the first nine equations of the full and restricted models range from .79 to .96. Each is significant beyond the .01 level. The shrinkage, or difference between the multiple correlation and the cross-validation correlation, is minimal, never exceeding .10.

Considerable shrinkage was noted in both the full and restricted models for the Flexibility and Originality dimensions of Activity 7. The attenuated cross-validation correlations of .56 and .48 in the full model and .59 and .62 in the restricted model certainly are at least of moderate value in view of the present state of the art; however, in comparison with previously stated results, they are somewhat disappointing. Hence, additional analyses were conducted, and a third model, the forced model, was generated.

As defined earlier, a forced model is one in which the researcher selects a potential set of predictors and forces them into the analysis before the remaining variables of the full set are allowed to enter. If the forced model is to differ from the full model, it must also be restricted.

Before considering the process of selecting the forced predictor variables, the rationale for using this type of model will be discussed. In multiple regression analysis, only the full model, after cross-validation, reflects ones ability to predict in what ever field is being studied. The results of restricted and forced models represent goals to be attained in future research, and each of those models must be applied to a new sample if their validity is to be evaluated. Thus, when working with models other than the full model, the researcher need not necessarily restrict his efforts to only the developmental sample. He must realize, however, that the non cross-validated results are tentative and contingent upon the assumptions, however implicit, utilized in his method of generating the restricted or forced model.

In this phase of the research, the forced predictor variables were selected by analyzing the correlations between the predictors and the criterion in <u>both</u> the developmental and the cross-validation sample. Only those predictors whose correlation with the criterion did not vary appreciably from one sample to the other were selected. The attenuated "cross-validation results" (Lord-Nicholson Formula) for the forced model in this study were very encouraging. Correlations of .77 and .70 for the Flexibility and Originality equations in Activity 7 were established. Of course, these results are tentative, and must be empirically tested in a new sample.

As previously indicated, in multivariate analysis it is often pointless to elaborate a hypothesis for each predictor, or to explain how each predictor met or failed to meet expectations. Since several of the predictors included in this study are "variables of opportunity," that is, variables which have been shown to be of value in earlier phases of the investigation strong relationships were not expected in all cases. Again, however, as Stone (1966) emphasized, category construction is considered the most crucial stage in content analysis. Thus, the correlations between the Flexibility and Originality dictionaries and the criterion as well as the step in which these category count variables entered the regression analyses for each activity will be considered. These results are given in Table 34. The Correlations of the dictionaries with the criteria are generally very high, ranging from .35 to .79 and significant beyond the .01 level. Furthermore, these predictors entered the regression analyses at extremely early stages. Thus, their usefulness in the prediction process is established.

TABLE 34

CORRELATION OF DICTIONARY WITH CRITERION, INCLUDING STEP ENTERED

Activity	Dimension	r	Step Entered
4	Flexibility	.79**	1
4	Originality	.64**	7
5	Flexibility	.77**	1
5	Originality	.61**	2
6	Originality	.57**	2
7	Flexibility	.67**	1
7	Originality	.35**	3

** Significant at .01 level

REFERENCES

Archambault, F. X. A Computerized Approach to Scoring Verbal Responses to the Torrance Tests of Creative Thinking. Unpublished Doctoral dissertation, The University of Connecticut, 1969.

Asher, J. J. Towards a Neo-Field Theory of Problem Solving. Journal of General Psychology, 1963, 68, 3-8.

Ausubel, D. P. The Psychology of Meaningful Verbal Learning. New York: Grune and Stratton, 1963.

Barcus, F. Communication Content: Analysis of the Research 1900-1958. Unpublished Doctoral dissertation, University of Illinois. 1959. Cited by P. Stone, D. Dunphy, M. Smith, and D. Ogilvie. The General Inquirer: A Computer Approach to Content Analysis. Cambridge, Mass.: M.I.T. Press, 1966.

Berelson, B. Content Analysis in Communication Research. Glencoe, Illinois: The Free Press, 1952.

Bish, Gertrude G. A Study of the Relationship of Intelligence, Achievement, Creativity, Anxiety and Confidence Among Intermediate Grade Pupils in a Suburban Area Elementary School. Unpublished Doctoral dissertation, George Washington University, 1964.

Budd, R. An Introduction to Content Analysis. Iowa City: University of Iowa School of Journalism, 1963.

Buel, W. B. Assessment of Creativity in a Research Setting. Journal of Applied Psychology, 1961, 45, 353-358.

Burkhart, R. C. and G. Bernheim. Object Question Test Manual. University Park, Pennsylvania: Department of Art Education Research, Pennsylvania State University, 1963. Mimeographed.

Cartwright, D. Analysis of Quantitative Material. In L. Festinger and D. Katz (Eds.), Research Methods in the Behavioral Sciences. New York: The Dryden Press, 1953.

Cicirelli, V. G. Form of the Relationship Between Creativity, IQ, and Academic Achievement. Journal of Experimental Psychology, 1965, 56, 303-308.

Erickson, G. The Predictive Validity of a Battery of Creative Thinking Tests and Peer Nominations Among University of Minnesota High School Seniors Seven Years Later. Unpublished Master's thesis, University of Minnesota, 1966. Cited by E. P. Torrance. Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, New Jersey: Personnel Press, 1966.

Fisher, G. A. The SCORTXT Program for the Analysis of Natural Language. Unpublished manuscript: Bureau of Educational Research, The University of Connecticut, 1968.

Fleming, E. S. and S. Weintraub. Attitudinal Rigidity as a Measure of Creativity in Gifted Children. Journal of Educational Psychology, 1962, 53, 81-85.

Garth, T. A Statistical Study of the Content of Newspapers. School and Society, 1916, 3, 140-144.

George, A. Quantitative and Qualitative Approaches to Content Analysis. In I. Pool (Ed.), Trends in Content Analysis. Urbana, Illinois: University of Illinois Press, 1959.

Ghiselin, B. The Creative Process and Its Relation to the Identification of Creative Talent. In C. W. Taylor (Ed.), The 1955 University of Utah Research Conference on the Identification of Creative Scientific Talent. Salt Lake City: University of Utah Press, 1956, pp. 195-203.

Goralski, P. S. Creativity: Student Teachers" Perceptions of Approaches to Classroom Teaching. Unpublished Doctoral dissertation, University of Minnesota, Mineapolis, 1964. Cited by E. P. Torrance. Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, New Jersey: Personnel Press, 1966.

Greene, J. F. Scoring Creativity Tests by Computer Simulation. Unpublished Doctoral dissertation, The University of Connecticut, 1970.

Griffin, D. F. Movement Presponses and Creativity. Journal of Consulting Psychology, 1958, 22, 134-136.

Guilford, J. P. Nature of Human Intelligence. New York: McGraw-Hill, 1967.

Hallman, R. J. The Commoness of Creativity. Educational Theory, 1963, 13, 123-127.

Hammer, E. F. Emotional Instability and Creativity. Perceptual Motor Skills, 1961, 12, 102-105.

Hiller, J. H. A Computer Search for Traits of Opinionation, Vagueness and Specificity-Distinction in Student Essays. Paper read at American Psychological Association, Washington, D. C., September, 1967.

Hoepfner, R. Review of Torrance Tests of Tests of Creativity. Journal of Educational Measurement, 1967, 4, 191-193.

Holstoi, O. Content Analysis Research in the Social Sciences. Paper read at IBM-Texas A&M Conference on Computers in Humanistic Research. College Station, Texas, 1966.

Houston, J. P. and S. A. Mednick. Creativity and the Need for Novelty. Journal of Abnormal Social Psychology, 1963, 66, 137-141.

Jacobson, T. C. and J. J. Asher. Validity of the Concept Constancy Measure of Creative Problem Solving. Journal of General Psychology, 1963, 68, 9-19.

Kracauer, S. The Challenge of Qualitative Content Analysis. Public Opinion Quarterly, 1952, 16, 631-642.

Kreuter, K. and Gretchen Kreuter. The Useful Genius: Intelligence, Creativity, and American Values. Saturday Review, October 17, 1964, pp. 64-67.

Kuder, G. F. Kuder Preference Record. Chicago: Science Research Associates, 1953.

Lasswell, H., D. Lerner and I. Pool. The Comparative Study of Symbols. Stanford: Stanford University Press, 1952.

Lieberman, J. N. Playfulness and Divergent Thinking: An Investigation of Their Relationship at the Kindergarten Level. Journal of Genetic Psychology, 1965, 107, 219-224.

Long, Barbara H., E. H. Henderson and R. C. Ziller. Self-Social Correlates of Originality in Children. Newark, Delaware: University of Newark, 1965, mimeo. Cited by E. P. Torrance. Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, New Jersey: Personnel Press, 1966.

Long, Barbara H. and E. H. Henderson. Opinion Formation and Creativity in Elementary Children. Newark, Delaware: University of Newark, 1964, mimeo. Cited by E. P. Torrance. Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, New Jersey: Personnel Press, 1966.

Lorge, I. and R. Thorndike. The Lorge Thorndike Intelligence Tests: Technical Manual. Boston: Houghton Mifflin, 1962.

Lowenfeld, V. Creativity and Art Education. School Arts, 1959, 59(2), 5-15.

Mackler, B. Creativity and Life Style. Unpublished Doctoral dissertation, University of Kansas, Lawrence, 1952. Cited by E. P. Torrance. Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, New Jersey: Personnel Press, 1966.

Marcotte, D. R. A Computerized Contingency Analysis of Content Graded Essays. Unpublished Doctoral dissertation. The University of Connecticut, 1969.

McClelland, D. C., et al. The Achievement Motive. New York: Appleton-Century-Croft, 1953.

McManus, J. F. Computer Evaluation of College History Examinations by Actuarial Strategies. Unpublished Doctoral dissertation, The University of Connecticut, 1968.

McNemar, Q. Psychological Statistics (3rd. ed.), New York: Wiley, 1962.

Mendelsohn, G. A. and Barbara B. Griswold. Differential use of incidental Stimuli in Problem Solving as a Function of Creativity. Journal of Abnormal Social Psychology, 1964, 68, 431-36.

Mosier, C. I. The Need and Means of Cross-Validation. Educational and Psychological Measurement, 1951, 11, 5-28.

Mueller, R. J. Can the Public School Foster Creativity? Saturday Review, December 19, 1964, pp. 48-49, 64.

Nelson, Joan F. The Construction of a Scale of Teacher Judgement of Pupil Creativity. Unpublished Master's thesis, University of Minnesota, Duluth, 1963. Cited by E. P. Torrance. Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, New Jersey: Personnel Press, 1966.

Osgood, C. The Representational Model and Relevant Research Methods. In I. Pool (Ed.), Trends in Content Analysis. Urbana, Illinois: University of Illinois Press, 1959.

Page, E. B. and D. H. Paulus. The Analysis of Essays by Computer. Final Report. Project No. 6-1318, Contract No. OEC-16-001318-1214. U S. Department of Health, Education and Welfare. Storrs, Conn.: The University of Connecticut, 1968.

Paulus, D. H. Are sub-tests of the Torrance Test independent? Paper presented at the annual meeting of the American Education Research Association, Minneapolis, Minn., March, 1970.

Paulus, D. H. and J. S. Renzulli. Computer Simulation of Human Ratings of Creativity. Storrs, Conn.: The University of Connecticut, 1968 (mimeo.).

Perry, Joyce M. Correlation of Teacher Prediction for Student Success Six Years Beyond Sixth Grade. Unpublished Doctoral dissertation, University of Illinois, Urbana, 1966.

Pool, I. (Ed.). Trends in Content Analysis. Urbana, Illinois: University of Illinois Press, 1963.

Renzulli, J. S. Problems and Issues in the Measurement of Creativity. Paper presented at the annula meeting of the American Education Research Association, Minneapolis, Minn., March, 1970.

Roget, P. (Ed.). Roget's International Thesaurus. New York: Crowell, 1962.

Sedelow, S., W. Sedelow and T. Ruggles. Some Parameters for Computational Stylistics. In J. Bessinger, S. Parrish, and H. Arader (Eds.), Literary Data Processing Conference Proceedings . New York: Modern Language Association, 1964.

Scule, R. Soule's Dictionary of English Synonyms. Boston: Little, Brown, & Co., 1966.

Speed, J. Do Newspapers Now Give the News. The Forum, 1893, 15, 705-711. Cited by P. Stone, D. Dunphy, M. Smith, & D. Ogilvie. The General Inquirer: A Computer Approach to Content Analysis. Cambridge, Mass.: M.I.T. Press, 1966.

Stein, M. I. A Translational Approach to Creativity. In C. W. Taylor (Ed.), The 1955 University of Utah Research Conference on the Identification of Creative Scientific Talent, Salt Lake City: University of Utah Press, 1956, pp. 195-203.

Stone, P., D. Dunphy, M. Smith and D. Ogilvie. The General Inquirer: A Computer Approach to Content Analysis. Cambridge, Mass.: M.I.T. Press, 1966.

Torrance, E. P. Guiding Creative Talent. Englewood Cliffs, New Jersey: Prentice Hall, 1962.

Torrance, E. P. Education and the Creative Potential. Minneapolis: University of Minnesota Press, 1963.

Torrance, E. P. Torrance Tests of Creative Thinking. Princeton, New Jersey: Personnel Press, 1966a.

Torrance, E. P. Torrance Tests of Creative Thinking: Direction Manual and Scoring Guide. Princeton, New Jersey: Personnel Press, 1966b.

Torrance, E. P. Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, New Jersey: Personnel Press, 1966c.

Weisberg, P. and K. J. Springer. Environmental Factors in Creative Thinking. Archives of General Psychiatry, 1961, 5, 554-564.

Willey, M. The Country Newspaper. Chapel Hill: University of North Carolina Press, 1926.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Yamamoto, K. Creative Writing and School Achievement. School and Society, 1963, 91, 307-308.

Yamamoto, K. Evaluation of Some Creativity Measures in a High School with Peer Nominations as Criteria. Journal of Psychology, 1964, 58, 285-293.

Yamamoto, K. A Study of the Relationship Between Creative Thinking Abilities of Fifth-Grade Teachers and Academic Achievement. Unpublished Doctoral dissertation, University of Minnesota, Minneapolis, 1964. Cited by E. P. Torrance. Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, New Jersey: Personnel Press, 1966.

Zieky, M. J. An Automated Aid to the Study of Literature the Application of Content Analytic and Statistical Techniques to Paradise Lost. Unpublished Doctoral dissertation, The University of Connecticut, 1968.